Journal of Hydrology 601 (2021) 126621

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Investigating spatial heterogeneity of the controls of surface water balance in the contiguous United States by considering anthropogenic factors

Zhiying Li *, Steven M. Quiring

Department of Geography, The Ohio State University, Columbus, OH, USA

ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief, with the assistance of Hanbo Yang, Associate Editor

Keywords:
Streamflow
Budyko
Spatial heterogeneity
Geographically weighted regression
Ordinary least squares regression

ABSTRACT

Understanding how precipitation is partitioned into evapotranspiration and streamflow is important for assessing water availability. In the Budyko framework, this partitioning is quantified through the ω parameter. Previous studies have modeled the physical representation of ω; however, the spatial heterogeneity of the relationship between ω and the variables that it represents has not been investigated. This study uses a geographically weighted regression model to identify spatial variations in the factors that control the water balance in 126 reference watersheds with minimal human disturbance and 765 non-reference watersheds in the continental United States. Results show that snowfall and forest coverage are important predictors of ω in the reference watersheds. Relative cumulative moisture surplus, dam storage, and developed land in riparian areas are important predictors in non-reference watersheds. Climate is a primary control of the relative importance of forest coverage. The importance of forest coverage is greater in arid watersheds than in humid watersheds. Snowfall is more important than forest coverage in the Northeast and Midwest. This study demonstrates that dam construction and urban sprawl have a significant impact in non-reference watersheds. Dam storage is the most important predictor in 21% of the non-reference watersheds, and riparian developed land is more important in 13% of the non-reference watersheds. Overall, there are statistically significant relationships between climatic, physiographic, and human-related factors and the ω parameter. The spatial variations in the relationship quantified in this study can help to improve regional watershed management.

1. Introduction

Understanding how precipitation is partitioned into evapotranspiration and streamflow is important for understanding global and regional water availability. Climate variability, watershed physiographic characteristics, and anthropogenic activities can substantially impact the surface water balance (Berghuijs et al., 2017; Gentine et al., 2012). Quantification of the relative importance of these factors is critical for improving water resources management and decision making. Process-based hydrological models are one method of quantifying the impacts of these factors (Dey and Mishra, 2017). However, it can be time-consuming to apply these models across many watersheds due to the labor-intensive model calibration process (Fatichi et al., 2016). In the last decade, a conceptual hydrological framework known as the Budyko framework has been successfully applied for many applications, such as quantifying runoff sensitivity (Berghuijs et al., 2017; Gudmundsson et al., 2016; Renner et al., 2012; Sankarasubramanian et al., 2001), unravelling the effects of climate and anthropogenic factors on streamflow (Jiang et al., 2015; Liang et al., 2015; Patterson et al., 2013; Wang and Hejazi, 2011; Wang et al., 2020), modeling streamflow and evapotranspiration (Abatzoglou and Ficklin, 2017; Chen et al., 2013; Fang et al., 2016; Nayak et al., 2020), and improving the calibration of global hydrological models (Greve et al., 2020). The advantages of the Budyko framework are multifold. First, it is physically based. The Budyko hypothesis relates the evaporative ratio (ET/P) to the aridity index (PET/P). The relationship is constrained by the energy limit when PET equals to ET and the water limit when ET equals to P. The second advantage is that it has a low computational cost, making it more efficient to apply over a large number of watersheds to examine spatial variability (Abatzoglou and Ficklin, 2017; Padrón et al., 2017; Xu et al., 2013). Lastly, it has fewer data requirements than process-based hydrological models, making it preferable for applications over longer time periods and larger spatial scales.

In the version of the Budyko framework that is known as Fu's equation (Fu, 1981), long-term streamflow is simulated using mean annual precipitation (P), mean annual potential evapotranspiration

* Corresponding author.
E-mail address: li.8254@osu.edu (Z. Li).

https://doi.org/10.1016/j.jhydrol.2021.126621
Received 1 April 2021; Received in revised form 30 May 2021; Accepted 24 June 2021
Available online 30 June 2021

(PET), and other factors that are represented by the shape parameter ω. The ω parameter is used to partition the water balance input (precipitation) and output (evapotranspiration and streamflow); however, it does not have a specific physical meaning (Abatzoglou and Ficklin, 2017). Previous studies have explored the meaning of ω and related it to climatic factors such as seasonality, storminess, and snowfall conditions, as well as watershed characteristics. For example, Potter et al. (2005) found that for a given aridity index, the ratio of mean annual E to P was larger when P and PET were in phase. When higher temperatures coincide with higher moisture surplus, this can increase evapotranspiration. Abatzoglou and Ficklin (2017) found a significant negative correlation between ω and the seasonal asynchronicity between P and PET. The importance of seasonality or synchronicity in P and PET was also emphasized by Milly (1994), Wolock and McCabe (1999), Shao et al. (2012), and Xing et al. (2018). Snowfall is another climate-related element that can affect the water balance. Using the Budyko framework, Berghuijs et al. (2014) found that streamflow is likely to decrease when precipitation shifts from snowfall to rainfall. This is consistent with a negative correlation between ω and the fraction of precipitation falling as snow shown in Abatzoglou and Ficklin (2017). Finally, storminess such as average storm depth and storm frequency has also been shown to influence ω (Donohue et al., 2012; Shao et al., 2012; Xing et al., 2018; Yang et al., 2009; Zhang et al., 2004).

Watershed characteristics such as vegetation, soil properties, and topography can also influence ω. Vegetation-related variables that can change ω include vegetation coverage and vegetation type (Abatzoglou and Ficklin, 2017; Chen et al., 2020; Donohue et al., 2007, 2010; Li et al., 2013; Shao et al., 2012; Sinha et al., 2019; Xu et al., 2013; Zhang et al., 2016). For example, Yang et al. (2008a) found a significant correlation between the curve parameter η (which is a shape parameter) of Choudhury's equation and the leaf area index. Forested watersheds have a higher ω than grassland-dominated watersheds because they typically have greater evapotranspiration (Zhang et al., 2004). Soil-related variables such as saturated hydraulic conductivity and available water holding capacity can also impact ω (Abatzoglou and Ficklin, 2017; Donohue et al., 2012; Yang et al., 2007). For example, the ratio of available water capacity to precipitation has been shown to have a statistically significant positive correlation with ω, as available plant water regulates the available water for Q (Abatzoglou and Ficklin, 2017). Watershed topography variables such as slope, aspect, relief ratio, and elevation have also been found to influence ω (Abatzoglou and Ficklin, 2017; Shao et al., 2012; Sinha et al., 2019; Xing et al., 2018; Xu et al., 2013; Yang et al., 2007, 2009). For example, slope is negatively correlated with ω because watersheds with steeper slopes tend to have higher runoff (Abatzoglou and Ficklin, 2017; Yang et al., 2007, 2009).

Human activities such as the modification of land use/land cover, urbanization, and irrigation have been observed to alter the hydrological cycle worldwide (Debbage and Shepherd, 2018; Destouni et al., 2013; Rodell et al., 2018). Although the non-parametric Budyko framework was initially developed for large-scale watersheds with minimal human interference, the parametric Budyko framework, such as the Fu's equation (Fu, 1981), can be applied to human-impacted watersheds (Liang et al., 2015; Patterson et al., 2013; Wang and Hejazi, 2011). Some recent studies have attempted to directly relate ω to factors such as irrigated area, cultivated land area, percentage of farmland, and population (Bai et al., 2020; Bao et al., 2019; Han et al., 2011; Jiang et al., 2015; Xing et al., 2018). A positive correlation between ω and irrigated areas was found in 96 watersheds in China (Xing et al., 2018). This is also found in sub-watersheds of the Yellow River Basin in China (Bao et al., 2019; Jiang et al., 2015). In the United States, human activities have been found to have greater impacts on streamflow than climate change in the High Plains and western U.S. (Wang and Hejazi, 2011). However, it is unclear which specific human-related factors are dominant and how they vary regionally. Abatzoglou and Ficklin (2017) modeled ω in HDCN and MOPEX watersheds in the continental U.S. and found that it was strongly influenced by climate factors, soil characteristics, and watershed topography. They pointed out that their model performed slightly worse in the MOPEX watersheds because the watersheds are influenced by irrigation and land use changes (Abatzoglou and Ficklin, 2017). The impacts of human activities on the hydrological cycle have been observed in many regions in the continental U.S. such as the Northeast (Hodgkins et al., 2019), Midwest (Kelly et al., 2016), and Southeast (Debbage and Shepherd, 2018). Therefore, it is important to identify how human disturbance influences ω in these watersheds.

Previous studies have evaluated how multiple factors control ω concurrently (Abatzoglou and Ficklin, 2017; Donohue et al., 2012; Li et al., 2013; Xing et al., 2018; Xu et al., 2013). For example, Xu et al. (2013) used a multiple linear regression model and a neural network model in the MOPEX watersheds and 32 global watersheds. They found that most of the variance in ω can be explained by geographic locations, Normalized Difference Vegetation Index (NDVI), slope, and elevation. At the continental scale, Abatzoglou and Ficklin (2017) used a generalized additive model to model ω in 382 watersheds in the continental U.S. They found that climate seasonality, snow fraction, the ratio of available water capacity to precipitation, and slope can explain 81.2% of the variability in ω. Xing et al. (2018) used a multivariate adaptive regression spline model to simulate η in Choudhury's equation in 96 watersheds in China. They found that the three most influential factors were average storm depth, vegetation coverage, and precipitation seasonality. They also found significant interaction effects from cultivated land, irrigation, drought, and precipitation variability. Despite excellent prior work modeling ω in different locations around the world, the spatial heterogeneity of the relationship between ω and the independent variables has not been studied. A negative correlation between ω and NDVI was found in a total of 211 HCDN watersheds in the U.S. (Abatzoglou and Ficklin, 2017) and 286 watersheds in China (Bai et al., 2020). However, a positive correlation between ω and NDVI was found in 224 MOPEX watersheds in the U.S. (Xu et al., 2013), 26 global watersheds (Li et al., 2013), and 96 watersheds in China (Xing et al., 2018). The contrasting relationship between ω and vegetation coverage was also shown by Yang et al. (2009) in China and Sinha et al. (2019) in India. Bao et al. (2019) found no significant correlation between ω and forest coverage in a watershed in China. The findings of these previous studies indicate that there may be spatial heterogeneity in the relationship between ω and explanatory variables. A generalized model may not be helpful for uncovering the relationship.

The lack of consideration of human-related factors and spatial heterogeneity of the relationship between ω and environmental variables may hinder our understanding of ω. Motivated by this research gap, three research questions are answered in this study: (1) Do the factors that influence ω differ between watersheds with and without human disturbance? (2) What are the most important factors that influence ω in watersheds with human disturbance? (3) How does the influence of these factors vary spatially? To answer the three questions, a geographically weighted regression model was used to investigate the spatial heterogeneity in the relationship between ω and 38 independent variables in 891 watersheds (126 reference and 765 non-reference watersheds) from 1950 to 2009 water years in the continental U.S.

2. Data and methods

2.1. Study area

This study evaluates 891 watersheds covering a wide range of scales, climate types, and topographic conditions in the contiguous U.S. The watersheds consist of 126 reference watersheds and 765 non-reference watersheds with continuous gauging records from 1950 to 2009 water years. The reference and non-reference watersheds are classified by the Geospatial Attributes of Gages for Evaluating Streamflow Version II (GAGES II) dataset (Falcone et al., 2010a). The reference watersheds have minimal human disturbance, while non-reference watersheds have

been disturbed by human activities such as development, irrigation, reservoirs, roadways, and fragmented land. The watersheds range in size from 480 km^2 to 49592 km^2. They also span a wide range of climate regions with a range of the aridity index from 0.3 to 5.3. The slope of these watersheds is also highly variable ranging from < 0.1% to 55.8% (Table 1). Watersheds smaller than 480 km^2 were not included in the study because of the possible inaccuracies of the representation of meteorological conditions from using the 4-km PRISM dataset (Abatzoglou and Ficklin, 2017). The reference watersheds are generally smaller than the non-reference watersheds and less variable in size (Table 1).

2.2. Budyko framework

The Budyko framework is a conceptual hydrological framework for estimating long-term water availability. Several parametric forms of the Budyko framework have been developed, and they have similar abilities in estimating streamflow and evapotranspiration (Jiang et al., 2015; Yang et al., 2008a, 2008b). Fu's version of the Budyko equation is used in this study (Fu, 1981). In Fu's equation, the evaporative ratio (ET/P) is a function of the aridity index (PET/P) and the ω parameter. The equation is expressed as:

Table 1
Summary statistics for mean annual P, PET, PET/P, Q, ω and the 38 independent variables in the reference and non-reference watersheds.

No.	Variables	Description	Reference (n = 126)			Non-reference (n = 765)		
			Min	Max	Mean ± SD	Min	Max	Mean ± SD
NA	Mean annual P (mm)	Precipitation	392.2	3072.0	1114.7 ± 492.5	330.8	2702.8	1001.7 ± 305.2
NA	Mean annual PET (mm)	Potential evapotranspiration	682.0	1844.3	1119.8 ± 223.9	675.1	2297.1	1151.3 ± 220.1
NA	Mean annual PET/P	Aridity index	0.3	4.7	1.2 ± 0.73	0.3	5.3	1.3 ± 0.6
NA	Mean annual Q (mm)	Streamflow	2.2	2574.6	502.5 ± 486.8	1.857	2131.9	380.1 ± 265.1
NA	ω	Budyko Fu's equation parameter	1.1	4.0	2.2 ± 0.5	1.3	4.1	2.2 ± 0.4
1	rCMS	Relative cumulative moisture surplus	<0.1	0.8	0.3 ± 0.2	<0.1	0.8	0.3 ± 0.1
2	SF	Snow fraction	0.1	0.7	18.1 ± 0.2	0.1	0.7	18.8 ± 15.2
3	AWC/P	Ratio of available water capacity to precipitation	0.1	0.6	0.2 ± 0.1	0.1	0.7	0.2 ± 0.1
4	PERMAVE (inch/h)	Average soil permeability	10.7	313.2	74.7 ± 59.4	11.0	309.1	73.5 ± 51.7
5	Slope (%)	Mean watershed slope	<0.1	55.8	11.6 ± 12.0	<0.1	42.3	8.7 ± 9.0
6	ASPECT_N	Aspect "northness". Ranges from −1 (facing north) to 1 (facing south).	−1.0	1.0	−0.3 ± 0.6	−1.0	1.0	−0.3 ± 0.6
7	TOPWET	Topographic wetness index	9.2	14.9	11.9 ± 1.2	9.7	15.0	12.2 ± 1.2
8	COMPACT	Watershed compactness ratio	0.3	2.8	1.5 ± 0.5	0.5	2.8	1.3 ± 0.4
9	ELEV_MEAN (m)	Mean watershed elevation	33.2	2578.2	614.3 ± 556.0	10.9	3318.1	642.0 ± 693.7
10	ELEV_MAX (m)	Max watershed elevation	57.0	3970.0	1052.2 ± 922.3	25.0	4408.0	1197.0 ± 1200.9
11	ELEV_MIN (m)	Min watershed elevation	4.0	1553.0	331.7 ± 331.5	−15.0	2525.0	362.4 ± 484.2
12	ELEV_MEDIAN (m)	Median watershed elevation	34.0	2561.0	608.9 ± 560.8	7.0	3353.0	627.9 ± 689.1
13	ELEV_STD (m)	Standard deviation of watershed elevation	6.0	529.7	122.5 ± 118.7	3.6	1100.8	140.1 ± 152.6
14	ELEV_SITE (m)	Elevation at gage location	6.0	1554.0	338.4 ± 334.1	1.0	2525.0	365.0 ± 477.6
15	RRMEAN	Relief ratio; (ELEV_MEAN - ELEV_MIN)/(ELEV_MAX - ELEV_MIN)	0.1	0.6	0.4 ± 0.1	0.1	0.7	0.4 ± 0.1
16	RRMEDIAN	Relief ratio; (ELEV_MEDIAN - ELEV_MIN)/(ELEV_MAX - ELEV_MIN)	0.1	0.7	0.4 ± 0.1	0.1	0.8	0.4 ± 0.1
17	AREA (km^2)	Watershed drainage area	485.4	25791.0	2149.5 ± 2912.4	480.4	49592.2	5908.2 ± 8485.6
18	FOREST (%)	Watershed percent "forest"	0.0	93.7	49.5 ± 29.7	0.0	92.3	39.3 ± 27.2
19	MAINS100_FOREST (%)	Mainstem 100 m buffer "forest"	0.1	92.1	41.0 ± 22.5	0.0	92.2	28.1 ± 19.1
20	RIP100_FOREST (%)	Riparian 100 m buffer "forest"	0.1	89.4	49.1 ± 26.6	0.0	90.0	38.5 ± 23.9
21	PCT_IRRIG_AG (%)	Percent of watershed in irrigated agriculture	0.0	11.2	0.5 ± 1.6	0.0	52.7	1.8 ± 4.8
22	PLANT (%)	Watershed percent "planted/cultivated"	0.0	93.0	24.9 ± 27.0	0.0	93.3	30.3 ± 28.8
23	MAINS100_PLANT	Mainstem 100 m buffer "planted/cultivated"	0.0	69.1	18.8 ± 19.2	0.0	90.0	18.6 ± 18.0
24	RIP100_PLANT	Riparian 100 m buffer "planted/cultivated"	0.0	88.6	21.2 ± 22.3	0.0	93.0	26.5 ± 25.4
25	NDAMS	Number of dams in watershed	0.0	475.0	21.6 ± 60.8	0.0	1740.0	82.1 ± 168.9
26	DDENS (No./100 km^2)	Dam density	0.0	14.9	1.0 ± 2.0	0.0	16.3	1.7 ± 2.1
27	STOR_NID (megaliter/km^2)	Maximum dam storage in watershed	0.0	226.9	8.9 ± 27.0	0.0	1348.2	71.7 ± 125.0
28	STOR_NOR (megaliter/km^2)	Nominal dam storage in watershed	0.0	120.7	4.7 ± 16.5	0.0	905.7	43.9 ± 93.7
29	MAJ_NDAMS	Number of "major" dams in watershed	0.0	12.0	1.0 ± 2.0	0.0	155.0	9.0 ± 18.0
30	MAJ_DDENS (No./100 km^2)	Major dam density	0.0	0.6	0.1 ± 0.1	0.0	9.0	0.2 ± 0.5
31	FRESHW_WD (megaliters/(year*km^2))	Freshwater withdrawal	0.5	1048.4	37.0 ± 118.3	0.6	1214.4	52.0 ± 85.1
32	DEV (%)	Watershed percent "developed"	0.0	9.7	3.6 ± 2.0	0.0	77.1	8.0 ± 8.4
33	PDEN (persons/km^2)	Population density in the watershed	0.0	62.0	7.7 ± 8.7	0.0	896.9	47.7 ± 98.2
34	MAINS100_DEV (%)	Mainstem 100 m buffer "developed"	0.0	25.9	5.5 ± 5.2	0.0	67.0	8.2 ± 7.9
35	RIP100_DEV (%)	Riparian 100 m buffer "developed"	0.0	16.6	4.0 ± 2.9	0.0	62.8	7.2 ± 6.5
36	ROADS_D (km/ km^2)	Road density	0.1	2.8	1.1 ± 0.4	0.1	6.1	1.6 ± 0.2
37	RD_STR_INTERS	Number of road/stream intersections	0.0	1.2	0.4 ± 0.2	0.0	1.8	0.6 ± 0.2
38	IMP (%)	Watershed percent impervious surfaces	0.0	1.7	0.5 ± 0.4	0.0	30.7	1.8 ± 2.9

$$\frac{ET}{P} = 1 + \frac{PET}{P} - \left[1 + \left(\frac{PET}{P}\right)^{\omega}\right]^{1/\omega} \quad (1)$$

where ET represents evapotranspiration, P represents precipitation, and PET represents potential evapotranspiration. All three variables are at the mean annual timescale. The ω controls partitioning of the water balance into input (P) and output (ET and streamflow Q). The ω represents the residual influence, other than PET and P, on water balance (Greve et al., 2020). The ω is calibrated by minimizing the mean absolute difference between observed ET/P and simulated ET/P by using ω values ranging from 1 to 9. The Budyko framework assumes that the water storage is negligible at long timescales. Therefore, the observed ET is calculated from P minus Q. The residual influence represented by ω is then modeled by independent variables in Section 2.3 to quantify the impacts of climatic, physiographic and anthropogenic factors on the water balance. Previous studies have applied the Budyko framework in reference and non-reference watersheds with varied study periods (Jiang et al., 2015; Wang and Hejazi, 2011; Xing et al., 2018; Zhang et al., 2019). The minimum study length was six years (Yang et al., 2007) and the maximum length was 100 years (Berghuijs et al., 2017). We believe our study period of 60 water years is long enough to assume that water storage is negligible ($\triangle S = 0$). Only the watersheds that obey the water balance assumption ($P > Q$ and $(P - Q) < PET$) were used in this study.

Precipitation data were obtained from the 4-km PRISM (Parameter-elevation Regression on Independent Slopes Model) AN81m dataset (PRISM Climate Group, 2014). Daily potential evapotranspiration was estimated using the Hargreaves and Samani equation (Hargreaves and Samani, 1982), which requires daily minimum and maximum temperature that were obtained from the 800-m TopoWx gridded dataset (Oyler et al., 2015). P and PET were aggregated for each watershed at the annual timescale for the 1950 to 2009 water years. Streamflow data were acquired from the United States Geological Survey.

2.3. Independent variables of ω

Based on an extensive review of previous studies, the ω parameter in the Budyko equation has been shown to be related to climatic, physiographic, and human-related variables. The frequently identified variables are climate seasonality (Abatzoglou and Ficklin, 2017; Shao et al., 2012; Xing et al., 2018; Zhang et al., 2004), vegetation coverage (Ning et al., 2019; Shao et al., 2012; Xu et al., 2013; Yang et al., 2009; Zhang et al., 2004), available water-holding capacity of the soil or hydraulic conductivity (Abatzoglou and Ficklin, 2017; Yang et al., 2009), slope (Abatzoglou and Ficklin, 2017; Sinha et al., 2019; Xu et al., 2013; Yang et al., 2009), and human-related variables such as irrigated area or agricultural area (Bai et al., 2020; Bao et al., 2019; Jiang et al., 2015; Oliveira et al., 2019; Xing et al., 2018). Therefore, in this study, a total of 38 independent variables that represent various climatic, physiographic, and human-related factors were considered in this study (Table 1).

There are two climate variables, relative cumulative moisture surplus (rCMS; Abatzoglou and Ficklin, 2017) and fraction of precipitation falling as snow (SF; Berghuijs et al., 2014). The rCMS represents climate seasonality, and it is calculated as $rCMS = \frac{\sum_{i=1(Jan.)}^{12(Dec.)} P_i - PET_i}{P_{ann}}$ when $P_i > PET_i$, where P_i and PET_i are monthly precipitation and potential evapotranspiration, P_{ann} is the annual precipitation. Monthly precipitation and mean temperature were used to calculate SF using the method from Dai (2008).

A total of 18 physiographic variables were evaluated, including: soil properties, vegetation, topography, morphology, and drainage area (Variables No. 3–18 in Table 1). The soil properties are the ratio of AWC to mean annual precipitation (AWC/P; Abatzoglou and Ficklin, 2017) and average soil permeability (PERMAVE). The soil permeability represents the ability of soil to transmit water and is related to the saturated hydraulic conductivity used in Yang et al. (2009). The vegetation variables considered in this study are: percentage of forest coverage (FOREST), mainstem 100 m buffer forest (MAINS100_FOREST), and riparian 100 m buffer forest (RIP100_FOREST) in the watershed. MAINS100_FOREST is the percent of forest coverage at the 100 m buffer at each side of the centerline of the main course of a river. RIP100_FOREST is the percent of forest coverage at the 100 m buffer at each side of the centerline of all streams in the watershed. MAINS100_FOREST and RIP100_FOREST have not been used in previous modeling studies, but they might be important since the spatial configuration of land use has been shown to influence streamflow characteristics (Debbage and Shepherd, 2018). The topographic variables include the mean, maximum, minimum, median, and standard deviation of the watershed elevation, elevation at the gage location, relief ratio for the watershed (Shao et al., 2012; Xing et al., 2018; Zhang et al., 2004), mean watershed slope (Abatzoglou and Ficklin, 2017; Xu et al., 2013; Yang et al., 2009), mean watershed aspect (Sinha et al., 2019; Xu et al., 2013), and mean topographic wetness index (Sinha et al., 2019; Xu et al., 2013). The topographic wetness index is calculated from the natural logarithm of the ratio of the watershed drainage area to the tangent of the slope gradient (Wolock and McCabe, 1995). The watershed compactness ratio is defined as the ratio of the drainage area and the square of the watershed perimeter. The compactness ratio is used because it provides a morphological representation of the watershed. Watershed shape is shown to influence the concentration time of streamflow (Jung et al., 2017).

There are also a total of 18 variables associated with human activities in the watershed. These include variables related to irrigation, planted/cultivated area, developed area, the number of dams in the watershed, dam density in the watershed, total dam storage in the watershed, freshwater withdrawal, and road density in the watershed (Variables No. 21–38 in Table 1). The freshwater withdrawal includes surface-water and groundwater (Maupin et al., 2014). These variables are chosen because land-use change, water use, and dam construction can directly impact water availability (Magilligan and Nislow, 2005; Wada et al., 2014; Wang and Hejazi, 2011).

The physiographic variables and human-related factors are obtained from the Geospatial Attributes of Gages for Evaluating Streamflow, version II (GAGES-II) dataset (Falcone et al., 2010a; Falcone, 2011). The GAGES-II dataset was published by the U.S. Geological Survey (USGS) and developed as part of a national effort to characterize stream gauges (Falcone et al., 2010a). These data were checked using standard USGS review procedures. Classification of reference and non-reference watersheds were determined using multiple sources of information, including a GIS-derived hydrologic disturbance index (Falcone et al., 2010b), local expert judgment, and a visual inspection of gauges using high-resolution imageries and topographic maps (Falcone et al., 2010a, 2010b). Watershed characteristics were compiled from commonly used quality-controlled national data sources such as the National Land Cover Database, National Inventory of Dams, 100-m National Elevation Dataset, and State Soil Geographic dataset. A potential source of uncertainty could be the impacts of the spatial resolution of the elevation dataset on the aggregated topographic variables at watershed levels. However, considering the size of the watersheds (>480 km^2) used in this study, a resolution of 100-m should be adequate. Another potential uncertainty may come from the accuracy of the land-use and land-cover classification. The primary source of low accuracy of the classification is from distinguishing the context of grass (Wickham et al., 2013). However, the land-use types considered in this study, forests, urban areas, and agricultural areas, have accuracies around 80% (Wickham et al., 2013). Watershed characteristics from the GAGES-II dataset have been successfully employed in many hydrological applications, such as analyzing runoff ratio (Chang et al., 2014), characterizing hydrologic change (Sawicz et al., 2014), and evaluating streamflow trends (Rice et al., 2015). Therefore, the GAGES-II dataset is considered to be reliable and appropriate for this study.

2.4. Ordinary least squares (OLS) regression

OLS is a global regression method that assumes the relationship between dependent and independent variables is spatially stationary (i.e., location independent). OLS assumes that the dependent variable and the residual of the model are normally distributed, and no collinearity exists between independent variables. Therefore, ω values were log-transformed to adjust the skewed gamma distribution (Greve et al., 2015). Pairwise Pearson correlation coefficients were calculated to check for collinearity between independent variables and tested at $p < 0.05$ level. The independent variables are z-standardized by differencing between the variable and the mean, then dividing by the standard deviation (Bring, 1994). The purpose of this standardization is to obtain standardized beta coefficients in the OLS model to compare variable importance. The lm function in R was used to model ω. The Variance Inflation Factor (VIF) was used to evaluate the collinearity of independent variables selected by the model. Variables that have VIF higher than a threshold of 5 were removed (Menard, 2002). Both forward and backward stepwise variable selections were conducted to obtain the optimal model whose independent variables are all statistically significant ($p < 0.05$) and that has the highest adjusted R^2 and the lowest mean absolute error (MAE). The independent variables selected by the stepwise regression model are used for further spatial analysis using Geographically Weighted Regression.

2.5. Geographically weighted regression (GWR)

The GWR model is a local spatial regression model and assumes that the relationship between dependent and independent variables is spatially nonstationary or location specific. GWR has been used in previous hydrological studies at a continental scale to explore the spatial controls of runoff variability (Chang et al., 2014), minimum river discharge (Rennermalm et al., 2012), and hydrologic responses to urbanization (Li et al., 2020). GWR is an expanded form of simple multiple regression equation and can be expressed as:

$$Y(x) = \alpha(u_i, v_i) + \sum_k \beta_i(u_i, v_i)x_k + e_i \quad (2)$$

where $Y(x)$ is a matrix of the dependent variable as a function of a matrix of independent variable × , α is the regression constant at the i_{th} location (u_i, v_i), (u_i, v_i) is the spatial location, β_i is the coefficient for the k_{th} independent variable at the i_{th} location (u_i, v_i), x_k is the k_{th} independent variable, and e_i is the residual at the i_{th} location. Therefore, the coefficients β_i's vary continuously as a function of the location. GWR achieves the spatially varying coefficients by fitting equations for observations falling within a fixed or adaptive bandwidth. The adaptive schemes can adjust the bandwidth according to the density of data. Similar to Li et al. (2020), an adaptive bandwidth was used in this study because the observations are not distributed evenly in the continental U. S. The bandwidth can be determined by cross-validation (CV) minimization or corrected Akaike Information Criterion (AICc) minimization. CV minimization is generally relevant to the accuracy of the model, while the AICc takes model complexity into account. Within each bandwidth, observations closer to a given location have greater weights in estimating the coefficients than observations further away. The weights matrices are created using a kernel estimator. Gaussian and bi-square kernels are commonly used kernel estimators (Chang et al., 2014; Mohammadinia et al., 2017; Yacim and Boshoff, 2019). The two methods of determining bandwidth (CV and AICc minimization) and the two kernel estimators (Gaussian and bi-square) were tested to determine the optimal GWR model. The model whose bandwidth is determined by CV minimization and weights matrices are determined by the bi-square kernel has the highest adjusted R^2 and lowest AIC in both reference and non-reference watersheds (Fig. S1). This model is used in this paper. The number of nearest neighbors in an adaptive bandwidth is 20 in reference watersheds and 31 in non-reference watersheds.

The GWR analysis was conducted using the GWmodel package in R (Gollini et al., 2013). The global Moran's I index was used to test the statistical significance of the spatial autocorrelation of the ω values and the residuals of the OLS and the GWR model. A Monte Carlo test was used to test the significance of the spatial variability of the model's coefficients. GWR coefficients may have local collinearity issues even if the collinearity does not show up in the OLS model (Wheeler and Tiefelsdorf, 2005). Thus, the VIF was used to evaluate local collinearity in GWR. The variables that have VIF > 5 were removed to ensure that variables are independent at a local scale, and the model form is the simplest with the fewest independent variables (Fig. S2). Although GWR allows the coefficients of the independent variables varied over space, the set of independent variables is usually constant across space. Since it is hypothesized that the water balance in reference watersheds has different controls than in non-reference watersheds, the two sets of watersheds are modeled separately. A flowchart showing an overview of the research design of this study is shown in Fig. 1.

3. Results

3.1. Spatial patterns of ω

High values of ω (>3) are found in the Great Plains and Florida, and low values of ω (<2) are in the Northeast and Northwest (Fig. 2). These spatial patterns are consistent for both reference and non-reference watersheds. Although the number of non-reference watersheds (765) is much larger than the reference watersheds (126), the range and variability of ω values in the two sets of watersheds are similar. The ω values in reference watersheds range from 1.1 to 4.0 with a mean and standard deviation of 2.2 ± 0.5. In non-reference watersheds, ω values range from 1.3 to 4.1 with a mean and standard deviation of 2.2 ± 0.4 (Table 1). A Kolmogorov-Smirnov test showed that distributions of ω values from the two sets of watersheds are not statistically significantly different ($p > 0.05$). Moran's I for ω is 0.19 ($p < 0.05$) in reference watersheds and 0.57 ($p < 0.05$) in non-reference watersheds. This indicates that the ω values of both sets of watersheds have statistically significant spatial autocorrelation.

3.2. Exploratory data analysis

In reference watersheds, ω values have the highest negative correlation with rCMS ($r = -0.74$) and the highest positive correlation with AWC/P ($r = 0.71$; Fig. 3). Other variables that are strongly correlated with ω ($|r| > 0.7$) are FOREST ($r = -0.74$) and RIP100_FOREST ($r = -0.73$). In non-reference watersheds, rCMS has the highest negative correlation with ω ($r = -0.64$), and the topographic wetness index (TOPWET) has the highest positive correlation ($r = 0.41$; Fig. 4). None of the variables has a $|r| > 0.7$ in the non-reference watersheds. Generally, variables that have statistically significant correlations with ω are similar in the reference and non-reference watersheds, except for a few variables such as average soil permeability (PERMAVE) and relief ratio (RRMEAN). The ω in the reference watersheds has a correlation of −0.21 with PERMAVE, but the correlation is not statistically significant in non-reference watersheds. In contrast, ω in the non-reference watersheds has a correlation of 0.17 with RRMEAN, but the relationship is not statistically significant in reference watersheds. The signs of the relationships are generally consistent between reference and non-reference watersheds. The ω is negatively correlated with rCMS, SF, Slope, elevation (mean, max, median, and std), FOREST, MAINS100_FOREST, RIP100_FOREST, and MAINS100_DEV. The ω is positively correlated with AWC/P, TOPWET, AREA, planted/cultivated lands (PLANT, MAINS100_PLANT, and RIP100_PLANT), irrigation lands (PCT_IRRIG_AG), number of dams (NDAMS), dam density (DDENS), and developed land (DEV). Although the sign of the correlation is consistent between both sets of watersheds, their magnitudes differ. For example, the correlation between ω and FOREST is −0.74 in reference watersheds and −0.58 in

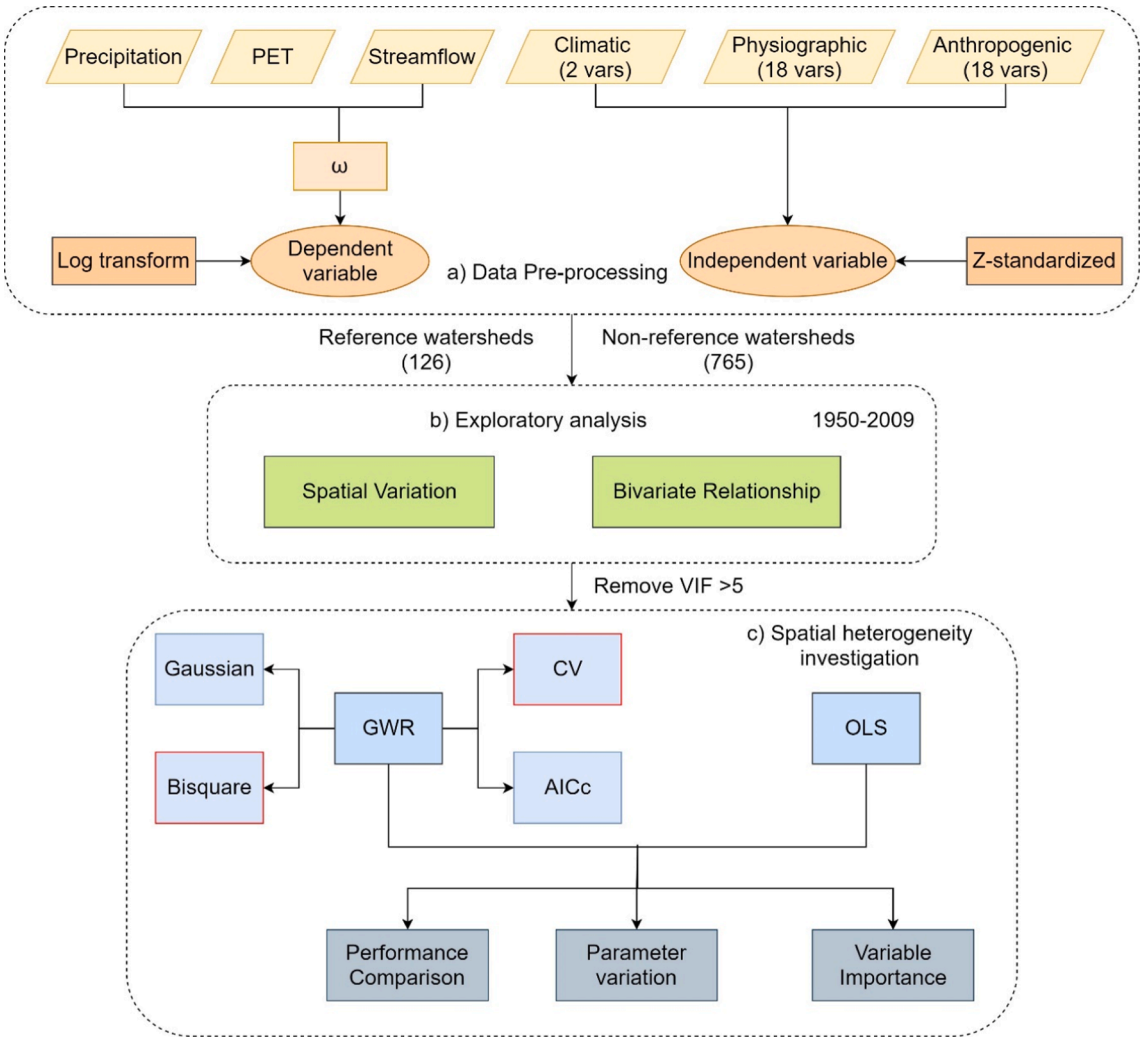


Fig. 1. Flowchart illustrating the research design of this study.

non-reference watersheds. This indicates the two types of watersheds may have somewhat different relationships and therefore should be modeled separately.

Statistically significant correlations also exist between some of the exploratory variables. Variables that belong to the same category have a high correlation, such as FOREST, MAINS100_FOREST, and RIP100_FOREST. The climatic variable rCMS has a positive correlation with slope ($r = 0.80$ in ref. and $r = 0.68$ in non-ref.); SF has a positive correlation with the median of watershed elevation ($r = 0.68$ in ref. and $r = 0.71$ in non-ref.); PLANT has a positive correlation with topographic wetness index ($r = 0.55$ in non-ref. and $r = 0.70$ in non-ref.). The collinearity was removed by retaining only one variable from each pair that are collinear. The variable was selected based on adjusted R^2, AICc, and MAE.

3.3. Spatial pattern of the controls and variable importance of ω

In reference watersheds, the OLS model shows that FOREST is the most important exploratory variable, followed by SF (Table 2). Both SF and FOREST have negative effects on ω. In non-reference watersheds, the important exploratory variables are different from the reference watersheds. Relative cumulative moisture surplus (rCMS) is the most important exploratory variable of ω, followed by dam storage (STOR_NOR) and riparian 100 m buffer developed land (RIP100_DEV; Table 2). This demonstrates that the controls of ω vary between watersheds with and without human disturbance. Human activities have a pronounced impact on ω in non-reference watersheds, especially those related to the construction of dams and urban sprawl. However, even in non-reference watersheds, ω represents more than just human activities. Climate variables such as rCMS also influence ω. Increases in STOR_NOR tend to increase ω, while increases in rCMS and RIP100_DEV tend to decrease ω. The OLS model for ω has an adjusted R^2 of 0.65 and an AICc value of −127.08 in reference watersheds, and an adjusted R^2 of 0.50 and an AICc value of −865.71 in non-reference watersheds. In comparison, the GWR model has an adjusted R^2 of 0.85 in reference watersheds and 0.80 in non-reference watersheds. The AICc value decreases to −194.29 in reference watersheds and −1367.57 in non-reference watersheds.

In GWR, the relationship between ω and independent variables varies across space. In both types of watersheds, a Monte Carlo (MC) test shows that all coefficients have statistically significant spatial non-stationarity ($p < 0.05$; Table 2). In reference watersheds, the coefficients of SF are negative in most watersheds in the eastern half of the country, but they are positive in the Northeast and several states such as North Dakota, Nebraska, and Texas (Fig. 5a). For FOREST, negative coefficients are found in most watersheds, especially in the Northern Great Plains (Fig. 5b). In the Northeast and Northwest, coefficients of FOREST are positive (Fig. 5b).

In non-reference watersheds, most watersheds have negative coefficients of rCMS, especially in Minnesota. In contrast, positive coefficients of rCMS are clustered in states such as Illinois, Mississippi, Florida, and Maine (Fig. 5c). For STOR_NOR, positive coefficients are found in most watersheds. They are distributed evenly across the

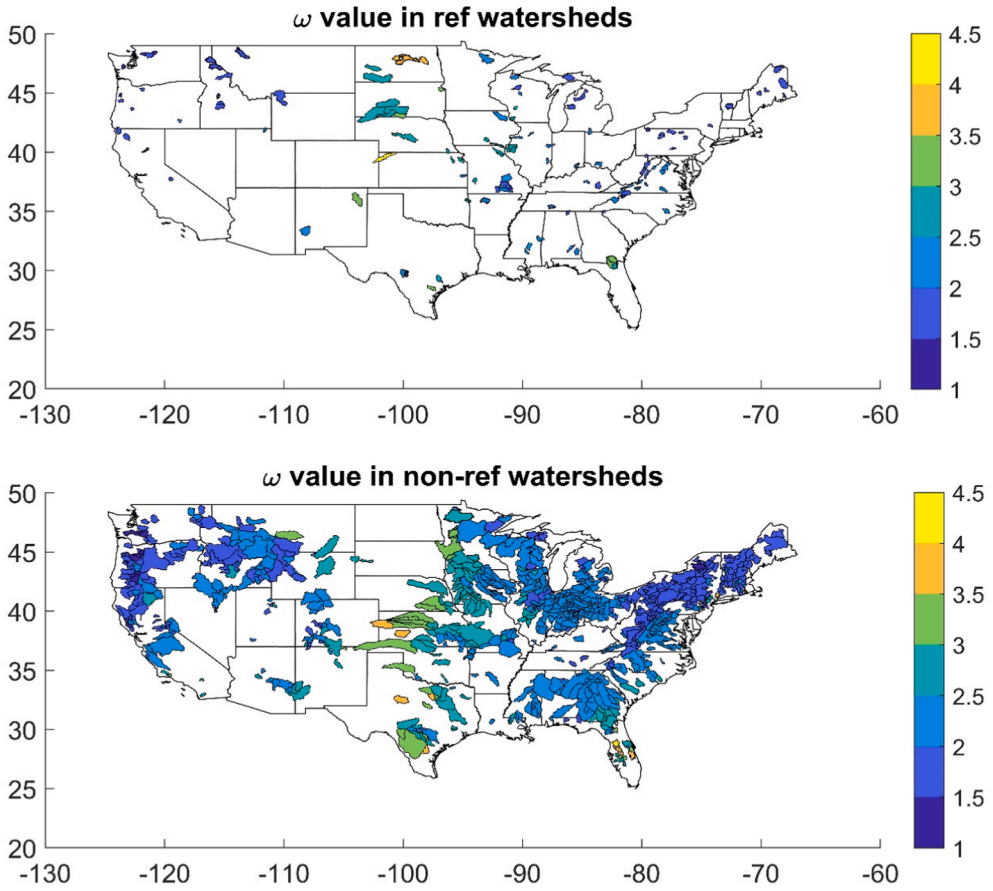


Fig. 2. Spatial variation of ω in reference watersheds (top; 126 watersheds) and non-reference watersheds (bottom; 765 watersheds) in the continental United States. The blue (yellow) colors indicate low (high) values of ω. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

country except in the Northwest, where most watersheds have negative coefficients of STOR_NOR (Fig. 5d). For RIP100_DEV, the western part mostly has positive coefficients, while in the eastern part, the number of watersheds with positive and negative coefficients is evenly split (Fig. 5e). About 80% of the watersheds (64 out of 80) in the Southwest have negative coefficients of RIP100_DEV, especially in Colorado.

The GWR model shows that SF is the most important variable in most reference watersheds, especially in the Northeast and Midwest. A total of 42 reference watersheds have FOREST as the most important variable. Most of them are along the west coast (Fig. 6a). A total of 506 non-reference watersheds have rCMS as the most important variable, and they are primarily located in the eastern half of the country (Fig. 6b). This is consistent with the OLS model in that rCMS is the most important variable. STOR_NOR is the most important variable in 159 watersheds (21%), and they are mostly located in the Midwest and Oregon. The remaining (13%) non-reference watersheds have RIP100_DEV as the most important variable, and they are in Colorado, Idaho, and parts of California.

3.4. Spatial pattern of local R^2

The local R^2 values of ω in reference watersheds are generally quite high and have a mean value of 0.87. The relatively low R^2 values (<0.6) are primarily found in Missouri (Fig. 7a). The modeled and observed ω have a correlation value of 0.93 ($p < 0.05$), and the mean absolute error is 0.12. As seen from the spatial pattern of the residuals (actual minus predicted), larger residuals tend to occur in watersheds in North Dakota, Nebraska, and California (Fig. 7b). Moran's Index for the OLS model is 0.05 ($p < 0.05$) and for the GWR model it is 0.02 ($p > 0.05$). This indicates that the residuals in the GWR model are randomly distributed (Table 2). The modeled Q closely matches the observed Q with an MAE of 26.38 mm and a MAPE of 9.59% (Fig. 8a).

In non-reference watersheds, the local R^2 values of ω are generally high and have a mean value of 0.79, which is slightly lower than the mean local R^2 in the reference watersheds. The locations with relatively low local R^2 values (<0.5) are primarily found in Indiana, Nevada, and parts of California (Fig. 7c). The modeled and observed ω have a correlation value of 0.92 ($p < 0.05$), and the mean absolute error is 0.11. The high values of residual of ω occur in several watersheds in California and Florida (Fig. 7d). Moran's Index for the OLS model is 0.49 ($p < 0.05$) and decreases drastically to 0.04 ($p > 0.05$) in GWR model. This indicates that the residuals in the GWR model do not exhibit spatial autocorrelation (Table 2). The modeled Q matches the observed Q with an MAE of 21.14 mm and a MAPE of 8.70% (Fig. 8b). While the mean value of R^2 values of ω in non-reference watersheds is slightly lower than in reference watersheds, the error in Q is also lower. This may be because the flow volume in reference watersheds is generally higher than in non-reference watersheds (mean value of 502.5 mm in ref vs 380.1 mm in non-ref). The same absolute difference of ω in humid and arid regions may not lead to the same absolute difference in streamflow.

4. Discussion

4.1. Spatial heterogeneity in the relationship between ω and predictors

According to the Monte Carlo test, all the independent variables in the GWR model have statistically significant spatial heterogeneity (Table 2). In the reference watersheds, coefficients of FOREST are

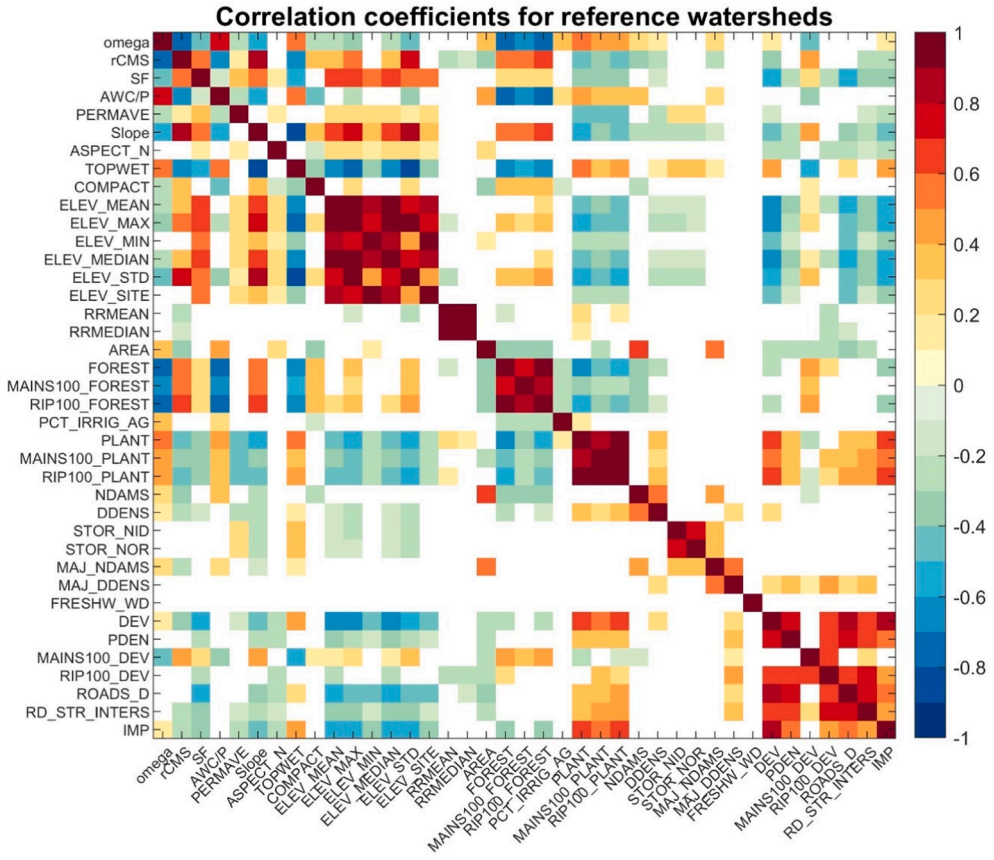


Fig. 3. Correlation matrix showing the correlation between ω and the independent variables in reference watersheds (n = 126). The red (blue) colors represent positive (negative) correlations. Correlations that are not statistically significant ($p > 0.05$) are shown in white. Descriptions of the variables are provided in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

negatively correlated with aridity index ($r = -0.79$, $p < 0.05$; Fig. 9). This indicates that climate is a major controlling factor of the effects of forest coverage on streamflow. Arid watersheds generally have higher absolute coefficients than humid watersheds. This is consistent with Zhang et al. (2016), who found that the Budyko shape parameter is more sensitive to vegetation change in dry watersheds. Watersheds with positive coefficients of FOREST are mostly located in regions where the aridity index < 1 (Fig. 9). This indicates that in humid regions, when forest coverage increases, ω and the evaporative ratio increase and this reduces Q/P. A positive correlation is also found in Li et al. (2013), Xu et al. (2013), and Xing et al. (2018). However, most watersheds have negative coefficients for the FOREST variable, especially when the aridity index is >1.2. Yang et al. (2009) also found a negative correlation between ω and vegetation coverage in 30 basins with an aridity index ranging from 1.12 to 2.60. The negative correlation suggests that decreases in ω and in E/P tend to be associated with an increase in forest coverage.

The physical mechanisms that are responsible for these relationships are due to the interactions between vegetation coverage, climate types, and water cycle (Gan et al., 2020). Changes in ω can be due to changes in PET/P or ET/P. When PET/P increases with an unchanged evaporative ratio (ET/P), ω decreases. This causes a horizontal shift in the Budyko curves (Fig. 10a). Forest coverage also decreases with increases in PET/P, as arid regions tend to have less forest (Fig. 10b). In this case, ω has a positive correlation with forest coverage. When ET/P increases and the evaporative ratio (PET/P) remains unchanged, ω increases. This causes a vertical shift in the Budyko curves (Fig. 10a). ET/P is negatively correlated with forest coverage when PET/P > 1 and it is positively correlated with forest coverage when PET/P < 1 (Fig. 10c). Therefore, in humid regions, when PET/P remains unchanged, ω tends to have a positive correlation with forest coverage. In contrast, in arid regions, ω tends to have a negative correlation with forest coverage. In other words, increases in forest coverage tend to increase ET/P and decrease Q/P in humid regions, while in arid regions, the opposite is true. It should be noted that the influences that are discussed here are specifically focused on the variation in relationships over space. Since arid watersheds are more widely dispersed in the contiguous United States than humid watersheds, the influence of forest coverage on water balance in arid watersheds may primarily be controlled by climate, while in humid watersheds, the influence may primarily be controlled by land-use types. Most reference watersheds in this study have PET/P > 1, and the changes in ω are mainly due to a vertical shift in the Budyko curves. Thus, the correlation between ω and forest coverage in most reference watersheds is negative, which is consistent with Bai et al. (2020) and Abatzoglou and Ficklin (2017). The decreases in forest coverage with increases in both the PET/P and ET/P was also found by Huo et al. (2021).

The spatial heterogeneity of coefficients of SF in the reference watersheds may be related to the threshold of the fraction of precipitation falling as snow. As seen from Fig. 9, most watersheds (85 out of 126) have negative coefficients of SF. This is consistent with Berghuis et al. (2014), who found a decrease in streamflow because of the precipitation shift from snow towards rain. The findings in Berghuis et al. (2014) do not apply to watersheds with marginal SF values (<15%). There are also 3 out of 97 watersheds used in Berghuis et al. (2014) that have opposite sensitivity of Q/P to SF. The watersheds with relatively high positive coefficients of SF (>0.1) in this study have SF values smaller than 17%. Regions with relatively low SF values may have higher chances that the increases in SF can lead to an increase in ω and E/P and a decrease in Q/P. The mechanisms that determine the influence of SF on streamflow

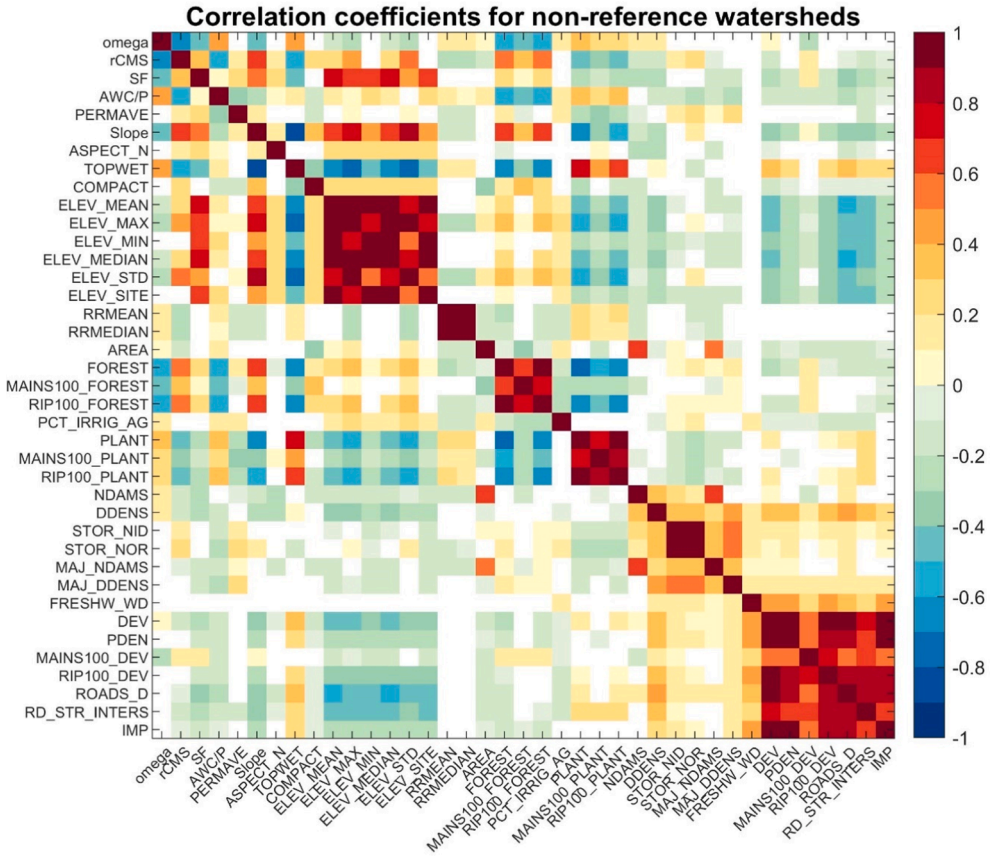


Fig. 4. Same as Fig. 3, but for non-reference watersheds (n = 765).

Table 2
Regression estimates and model goodness of fit of OLS and GWR. Moran's I represents the spatial autocorrelation of the model residuals (* indicates significance at 0.05 level). Monte Carlo (MC) test shows the p-value of spatial heterogeneity. A p-value < 0.05 indicates significant spatial heterogeneity.

Reference watersheds (n = 126)							
	OLS Estimate	GWR					
		Min	Q1	Median	Q3	Max	MC test
(Intercept)	0.80	0.32	0.57	0.67	0.82	1.31	<0.01
SF	−0.09	−0.54	−0.26	−0.11	<0.01	0.60	<0.01
FOREST	−0.14	−0.48	−0.11	−0.03	0.01	0.08	<0.01
AICc	−127.41						−194.29
Adj. R^2	0.65						0.85
Moran's I	0.05*						0.02
Non-reference watersheds (n = 765)							
	OLS Estimate	GWR					
		Min	Q1	Median	Q3	Max	MC test
(Intercept)	0.78	0.34	0.64	0.75	0.83	1.42	<0.01
rCMS	−0.17	−0.82	−0.25	−0.14	−0.05	1.08	<0.01
STOR_NOR	0.06	−0.94	−0.04	0.02	0.08	0.99	0.01
RIP100_DEV	−0.02	−0.50	−0.02	<0.01	0.03	0.52	<0.01
AICc	−865.71						−1367.57
Adj. R^2	0.50						0.80
Moran's I	0.49**						0.04

may be a combination of vegetation and topography and vary from watershed to watershed (Berghuijs et al., 2014). The decreases in streamflow as a result of a reduction of SF may be associated with increases in evapotranspiration due to declines in snow cover (Milly and Dunne, 2020) or increases in atmospheric demand for water in a warming climate (Neto et al., 2020).

In non-reference watersheds, coefficients of rCMS are negative in most watersheds except in the regions around Wisconsin, Illinois, and Indiana. This may be related to the climate region. As seen from Fig. 11, watersheds that have relatively high positive coefficients of rCMS are clustered around PET/P = 1. Results in the transitional areas where PET/P = 1 may have greater uncertainties. This is because the calculation of the rCMS is based on the difference between PET and P, and the two variables have similar values in the transitional areas. This may explain the relatively low local R^2 in regions in Midwest around Great Lakes. The negative coefficients of rCMS indicate that higher values of rCMS influence water balance by lowering ω values and increasing Q/P. This is because high values of rCMS represent an out-of-phase

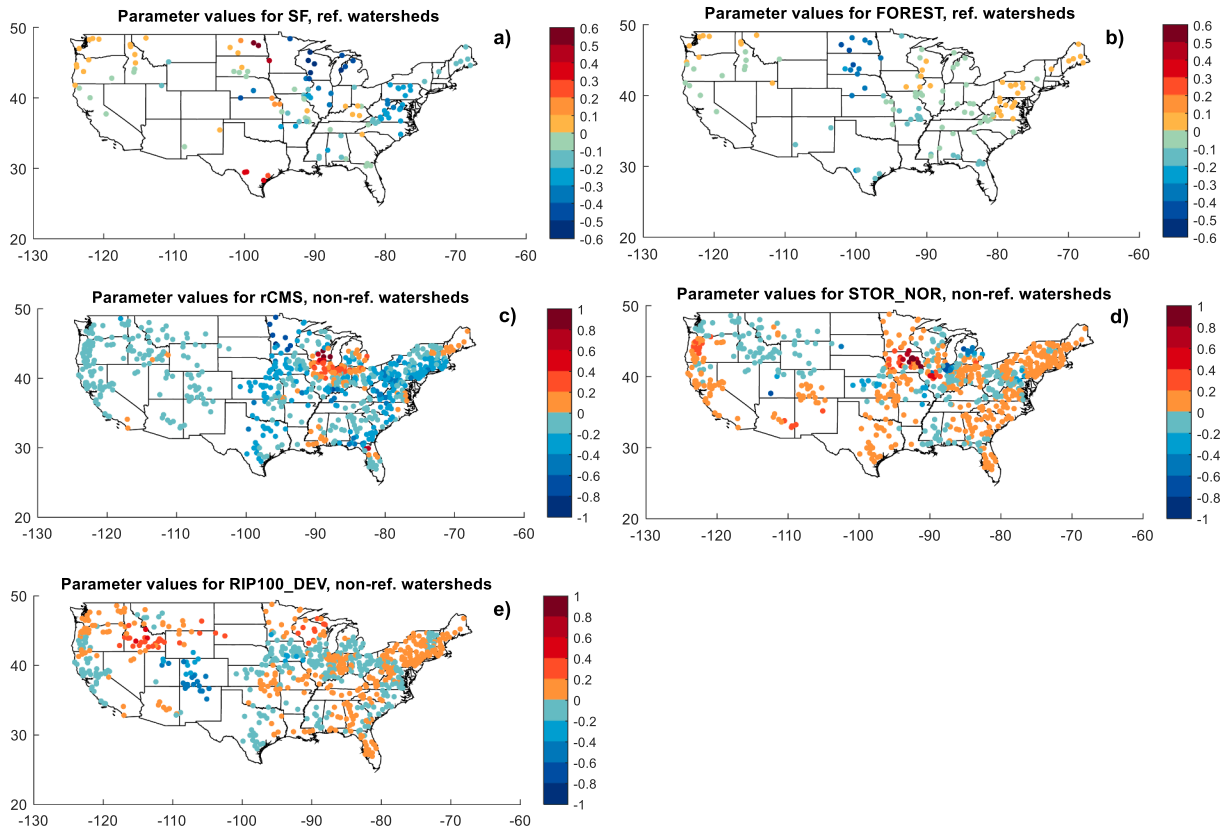


Fig. 5. Spatial variation of coefficients of GWR model of SF (a), FOREST (b) in the reference watersheds, and rCMS (c), STOR_NOR (d), RIP100_DEV (e) in non-reference watersheds. The red (blue) colors indicate positive (negative) coefficients. Note that the scale used in the reference watersheds (a and b; −0.6 to 0.6) is different from the scale used in the non-reference watersheds (c to e; −1 to 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

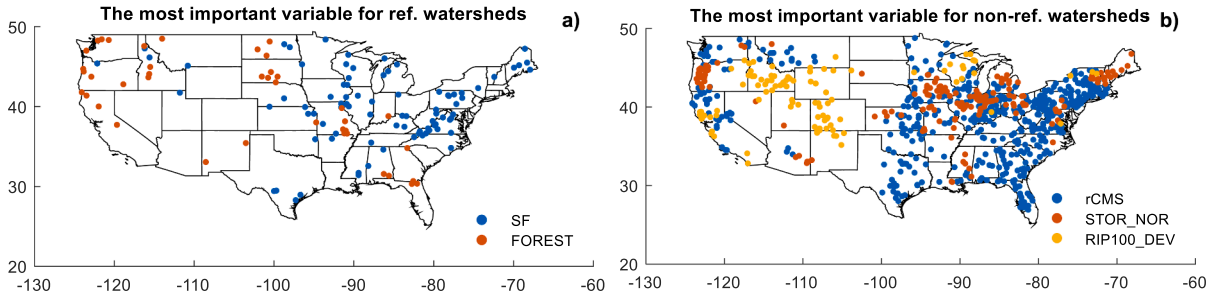


Fig. 6. Spatial distribution of the most important variable for reference watersheds (a) and non-reference watersheds (b). The most important variable is identified as the one with the highest absolute value of standardized beta coefficients.

relationship between PET and P (i.e., P is higher while PET is lower). In this case, a greater fraction of P can be converted to Q instead of ET (Wolock and McCabe, 1999).

STOR_NOR and RIP100_DEV are two human-related factors represented by ω in the GWR model in non-reference watersheds. The spatial distributions of their coefficients are more varied, and the controlling factors require further investigation. About 64% of the non-reference watersheds have positive coefficients of STOR_NOR. This suggests that an increase in dam storage can increase ω and E/P and decrease Q/P. This is consistent with the discussion in Wang and Hejazi (2011) that evaporation can be enhanced due to increases in the surface area of water bodies. For RIP100_DEV, about half of the watersheds have negative coefficients. The negative coefficients mean that increases in developed land in the riparian zone can decrease ω and E/P, and increase Q/P. The impacts of urban sprawl on increases in streamflow is also observed in previous studies in regions that have negative coefficients of RIP100_DEV, such as in Texas (Olivera and DeFee, 2007), the Charlanta Megaregion (Debbage and Shepherd, 2018), and Colorado (Bhaskar et al., 2020). The increasing impacts on streamflow are likely due to increases in impervious areas and decreases in infiltration. The positive coefficients may be associated with other processes during urbanization, such as cross-basin transfers of public and/or sewer water (Oudin et al., 2018). Different signs of the coefficients of developed land are also observed in Li et al. (2020), using the GWR model in exploring the impacts of urbanization on hydrology. Elevation may be a controlling factor of the importance of RIP100_DEV. As shown in the relationship between absolute values of the coefficients of RIP100_DEV and the mean elevation in the watershed, the importance of RIP100_DEV increases with elevation ($r = 0.71$; Fig. 10). This indicates that there may be a confounding effect between developed land and topography. Urban expansion in high-elevation regions may have a greater impact on runoff generation than in low-lying regions. Debbage and Shepherd (2018) also

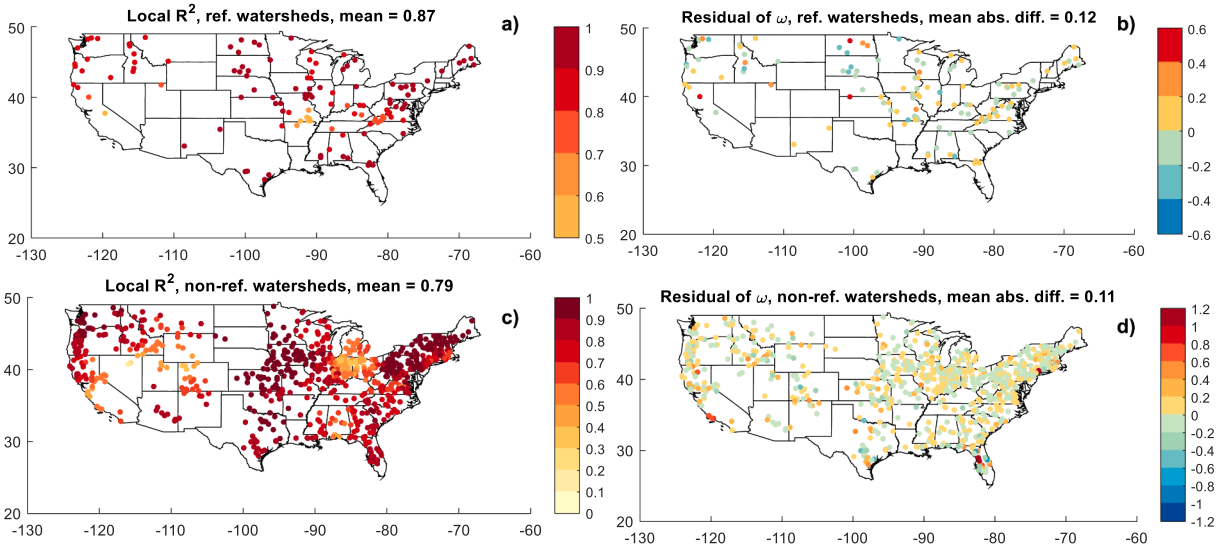


Fig. 7. Spatial pattern of local R^2 values and residuals of omega in reference watersheds (a and b) and non-reference watersheds (c and d). For local R^2 values, the darkest colors represent the highest value. For residuals of omega, the red (blue) colors represent positive (negative) residuals, which are calculated using the actual values minus the modeled values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

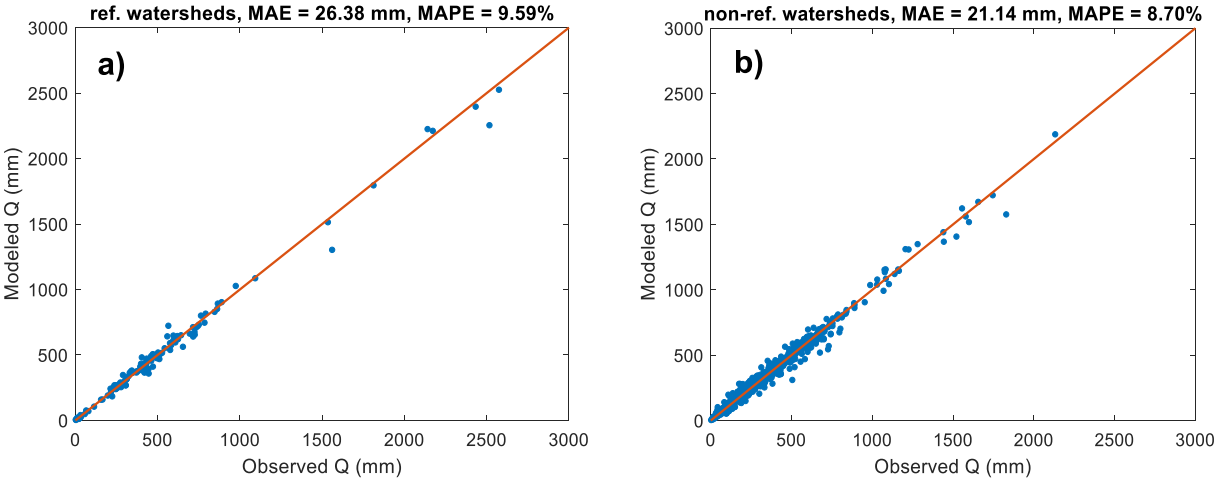


Fig. 8. Scatter plot of modeled and observed Q in reference watersheds (n = 126; a) and non-reference watersheds (n = 765; b).

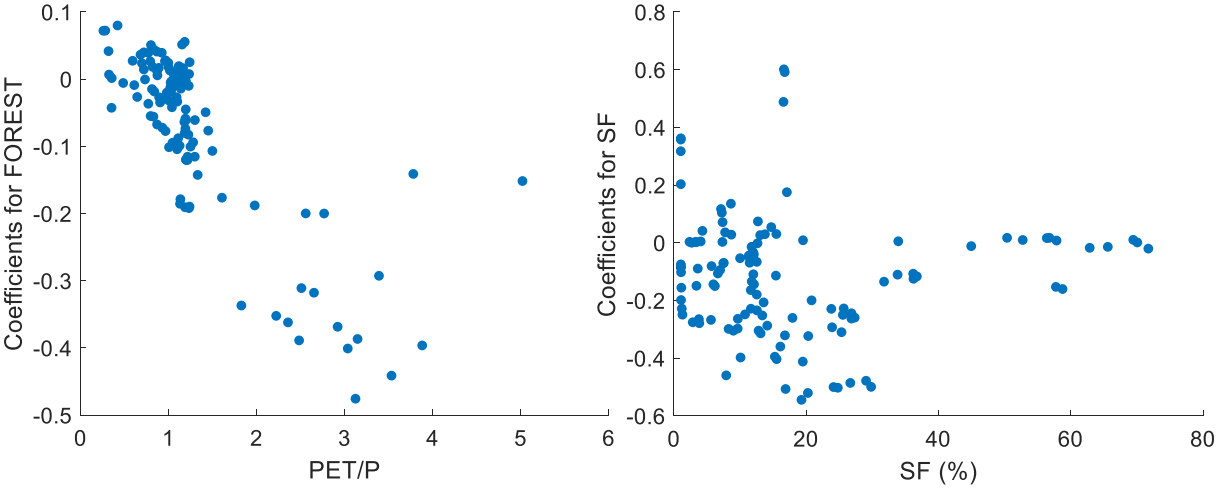
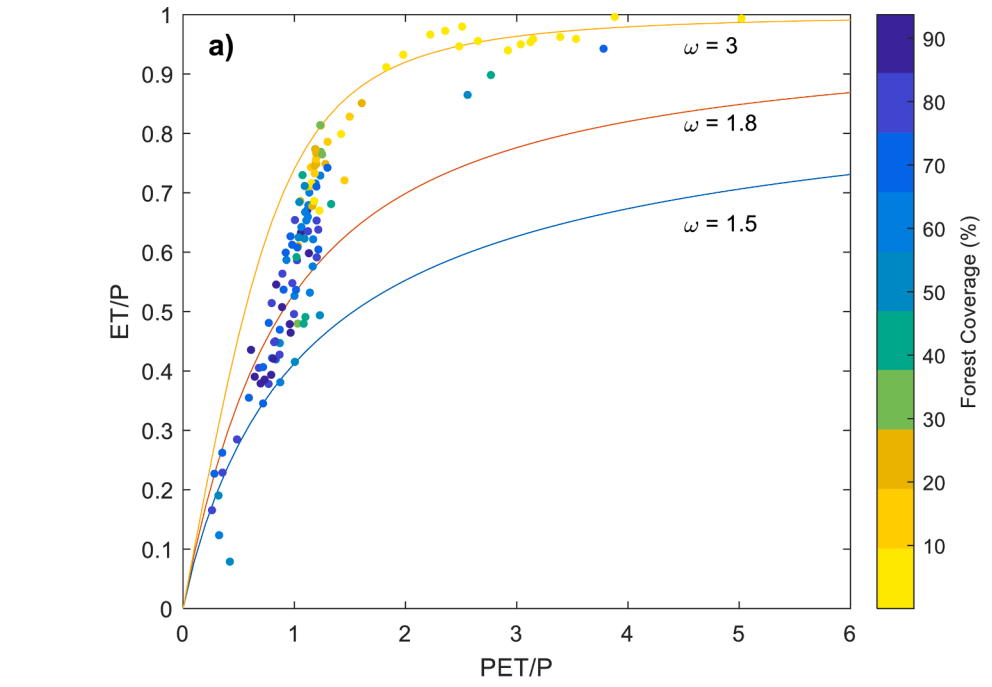
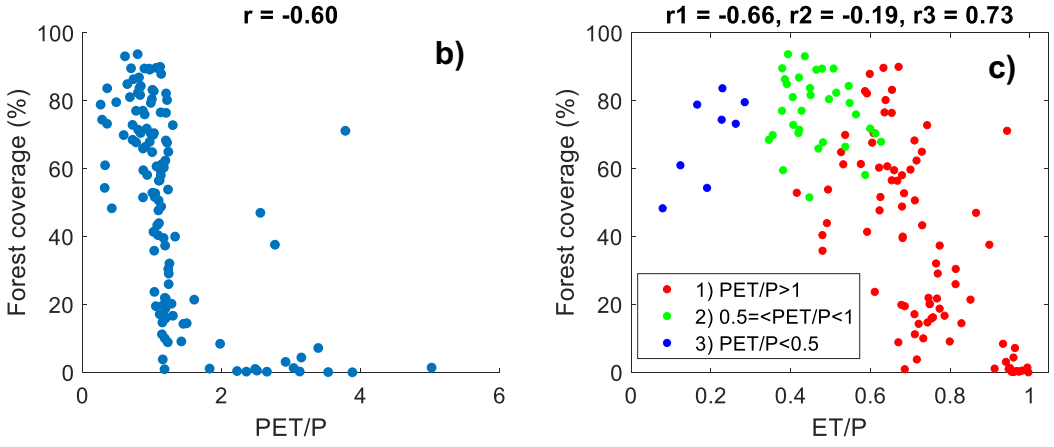


Fig. 9. Scatter plot of coefficients of FOREST against PET/P (left) and coefficients of SF against SF (right) in reference watersheds.





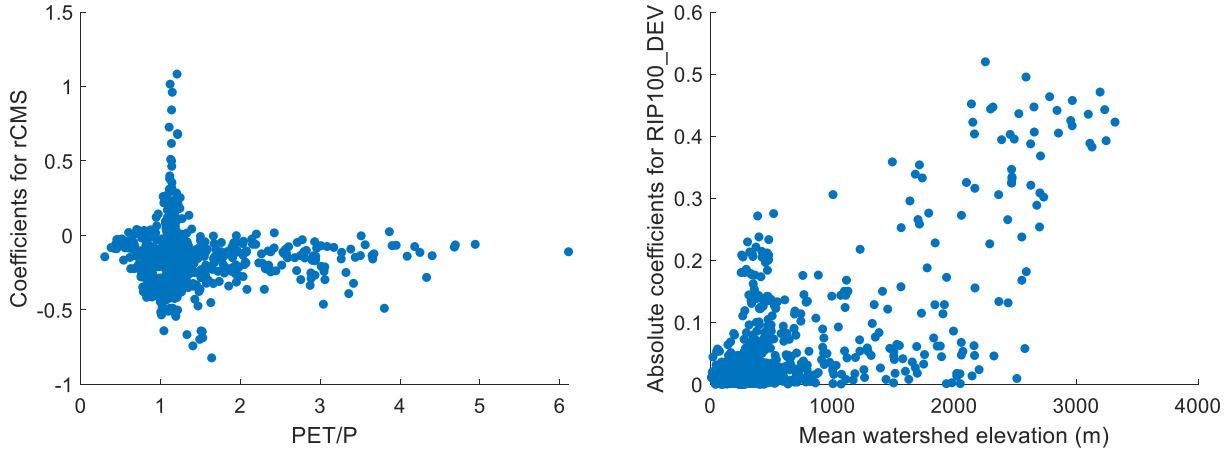


Fig. 10. a) Distribution of the forest coverage of reference watersheds (n = 126) on the Budyko curves (ω = 1.5, 1.8, 3). The yellow (blue) colors indicate lower (higher) values of the forest coverage. b) Scatterplot between PET/P and forest coverage in reference watersheds. "r" represents Pearson correlation coefficient. c) Group scatterplot between ET/P and forest coverage in reference watersheds. Red (blue) colors represent watersheds with PET/P > 1 (<0.5), and the green colors represent watersheds with 0.5 =< PET/P < 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 11. Scatter plot of coefficients of rCMS against PET/P (left) and absolute coefficients of RIP100_DEV against mean watershed elevation (right) in non-reference watersheds.

found that clustering impervious surfaces in source areas of the watershed played an important role in increasing high flow frequency.

4.2. Bivariate relationship

Low values of ω characterize conditions where precipitation is quickly converted to runoff, while high values of ω characterize a high ratio of ET to P (Dingman, 2015). This section discusses the possible physical mechanisms that are responsible for the relationship between ω and the variables that are not discussed in Section 4.1.

Physiographic variables that have statistically significant negative correlations with ω include slope, elevation, and watershed compactness ratio. The slope has a negative correlation with ω because P is quickly converted to Q in steeply sloping watersheds, leading to a smaller value of ω. This is consistent with Abatzoglou and Ficklin (2017) and Yang et al. (2007). Elevation has a similar physical mechanism influencing ω, and the negative correlation we found is consistent with the results of Xu et al. (2013). The watershed compactness ratio represents the watershed shape, and its bivariate relationship with ω was not explored in previous studies. A high value of the compactness ratio means a more compact and regular shape of a watershed, such as a circle or a rectangle (Bogaert et al., 2000), which leads to a shorter concentration time of flow, a faster conversion from P to Q, and a smaller value of ω.

Physiographic variables that are positively correlated with ω include AWC/P, topographic wetness index, drainage area, and relief ratio. A higher ratio of AWC/P characterizes greater subsurface storage, which reduces the conversion from P to Q and increases the value of ω (Abatzoglou and Ficklin, 2017; Xing et al., 2018; Yang et al., 2007). The topographic wetness index measures the long-term soil moisture availability (Kopecký and Čížková, 2010) and thus as it increases, so does potential ET and the value of ω. Drainage area has a similar physical mechanism influencing ω, and a positive correlation was also found in Xu et al. (2013). High values of the relief ratio characterize broad and level surfaces with several depressions (Pike and Wilson, 1971), reducing Q potential and increasing the ω value. A negative correlation between relief ratio and ω was found in Xing et al. (2018), but their relief ratio was calculated as the ratio of the difference between the maximum and minimum elevations and the longest flow path. High values of the relief ratio in Xing et al. (2018) represent a larger elevation change per unit length of the river, which is the opposite topographic conditions from the level surfaces indicated by the high relief ratio used in this study. In this study, relief ratio is calculated as the ratio of the difference between mean and minimum elevation and the difference between maximum and minimum elevation (Table 1). Therefore, our results agree with theirs.

Positive correlations were found between most human-related variables and ω. For example, ω is positively correlated with planted/cultivated land, which is consistent with previous studies (Bao et al., 2019; Han et al., 2011; Jiang et al., 2015). The physical mechanism may be because the irrigation in the planted/cultivated land leads to the enhanced ET (Wang and Hejazi, 2011).

4.3. Implications, limitations, and future studies

This study has implications for integrated watershed management at the local scale. Our study found that in reference watersheds, the impacts of forest coverage is more important in dry regions than in wet regions. This indicates that changes in forest coverage in water-limited regions should be undertaken with caution because they can have a dramatic impact on ET and Q. In non-reference watersheds, RIP100_DEV is the most important human-related variable. This indicates that these watersheds are particularly sensitive to development in riparian areas. Local water resources managers should carefully consider the spatial location and type of development that takes place in the watersheds, especially development in proximity to streams. In addition, the importance of RIP100_DEV increases with elevation. Watershed managers and city planners should be cautious about expanding urban areas in high elevation areas. Expansion of impervious surfaces in regions with high elevation can lead to a shorter concentration time and greater streamflow. Since this study identifies the most important variable in each watershed, this can be used to identify where water resources management efforts should be focused. Overall, we demonstrated that there are statistically significantly variations in terms of which variable has the strongest influence on the water balance in the contiguous United States. This indicates that effective water resources management should consider the local climate, land-use types, and topographic conditions and that a one-size-fits-all approach to water management will not be effective.

Although this study considers a comprehensive list of climatic, physiographic, and human-related independent variables, not all possible independent variables are included. For example, the average storm depth was found to be important for explaining ω in previous studies (Shao et al., 2012; Xing et al., 2018; Zhang et al., 2004), but it was not considered in this study. Even though it is not included as an independent variable, the performance of the current GWR model (adjusted $R^2 = 0.85$ for reference and 0.80 for non-reference watersheds) is reasonable for analyzing the spatial heterogeneity in the controls of streamflow. Storm depth is also likely correlated with other climate variables such as rCMS. In addition, climate seasonality can be represented by multiple variables such as Milly's index (Milly, 1994), the seasonality index developed by Walsh and Lawler (1981), precipitation seasonality developed by Markham (1970), and relative cumulative moisture surplus (Wolock and McCabe, 1999). This study only uses rCMS because of its simple calculation and the strong correlation with ω reported by Abatzoglou and Ficklin (2017). Future studies could consider other measures of seasonality.

GWR assumes no collinearity among the independent variables in the model (Wheeler, 2019). Therefore, this study removed variables where the local VIF > 5. For example, the original best global multiple regression model of ω in reference watersheds is $\omega = \exp(-0.09*\text{rCMS} + (-0.08)*\text{SF} + 0.08*\text{ELEV\_MEDIAN} + (-0.10)*\text{FOREST})$. The adjusted R^2 is 0.76, and all the variables have VIF < 5. However, if all four variables are used in the GWR model, many local observations have VIF $>$ 5. This is because local collinearity effects can be present even when no global collinearity is detected (Wheeler, 2007). The absence of collinearity in the global regression model is not a reliable indicator for the absence of local collinearity (Wheeler, 2007). To deal with this issue, we removed variables that lead to local VIF > 5 in both GWR and the counterpart OLS model. The final GWR model has all the observations with local VIF < 5 (Fig. S2) and the best set of model performance measures (R^2, AIC, and MAE). Removing variables is a simple and effective way to address local collinearity. This study provides reasonably good performance of the GWR model and uncovers the spatial heterogeneity of the relationship between ω and dominant independent variables. Future studies may consider using penalized versions of GWR, such as the geographically weighted ridge regression (GWRR), to retain all the variables in the global regression model while properly addressing local collinearity issues.

The GWR model can accurately predict ω values, and it has an adjusted R^2 of 0.85 in reference watersheds. This is higher than the R^2 of 0.812 obtained from a generalized additive model in 211 reference watersheds in the continental U.S. (Abatzoglou and Ficklin, 2017). Abatzoglou and Ficklin (2017) also validated their model in an independent set of watersheds and had an R^2 of 0.65. This study does not validate the GWR model because it focuses on exploring the spatial heterogeneity of the relationship between ω and the environmental predictors. Future studies can compare the accuracy of the GWR model with other modeling methods using validation watersheds for prediction purposes. Although the GWR model has good model accuracy, there are regions with relatively low values of local R^2, such as non-reference watersheds in Indiana, Ohio, and Michigan. The regions may have important independent variables in explaining ω other than rCMS,

STOR_NOR, and RIP100_DEV identified in this study, such as irrigation agriculture or planted/cultivated land percentage (Wang and Hejazi, 2011). Future studies can conduct separate modeling at regional scales to improve the model accuracy.

5. Conclusions

This study improves our understanding of the controls of surface water balance by uncovering the spatial heterogeneity of the relationship between Budyko shape parameter ω and environmental variables. The spatial-varying relationships help to clarify the relative importance of forest coverage as a function of climate. Human-related variables were also considered. This is an improvement over previous studies that only considered climatic and physiographic variables. A geographically weighted regression model was used to identify the set of variables that best explain the spatial variability of ω in 126 reference and 765 non-reference watersheds in the contiguous U.S. An ordinary least squares model was used to uncover the global relationship between the ω and the same set of environmental predictors. Our new findings are as follows:

(1) The representations of ω are different in reference and non-reference watersheds. In reference watersheds, ω is represented by precipitation falling as snow and forest coverage. In non-reference watersheds, ω is represented by relative cumulative moisture surplus, normal dam storage, and riparian developed land. The GWR model explains 85% of the variability in ω in reference watersheds and 80% of the variability ω in non-reference watersheds.
(2) In non-reference watersheds, human disturbance primarily consists of dam construction and urbanization. Dam storage is the most important variable in 21% of non-reference watersheds. Most of them are in parts of the Midwest and Oregon. Riparian developed land is more important in 13% of non-reference watersheds, and they are mostly in Colorado, Idaho, and parts of California.
(3) There are statistically significant spatial variations in the relationship between ω and the independent variables ($p < 0.05$). In reference watersheds, climate is a dominant factor controlling the spatial heterogeneity of coefficients of forest coverage. The relative importance of forest coverage is higher in dry watersheds than in humid watersheds. In non-reference watersheds, the spatial patterns of coefficients of dam storage and riparian developed land are more varied. The impact of riparian developed land on streamflow increases with elevation ($r = 0.71$).

The spatially varying importance of environmental and anthropogenic predictors identified in this study can be used for guiding local water resources management. Watershed managers and city planners should be cautious about changing forest coverage in reference watersheds in dry regions and expanding riparian developed areas in non-reference watersheds in high-elevation areas. This study identifies the most important variables, other than precipitation, that control local water availability. National watershed management strategies should be adapted at the local scale to reflect these spatially varying relationships.

CRediT authorship contribution statement

Zhiying Li: Visualization. **Steven M. Quiring:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by National Science Foundation Doctoral Dissertation Research Improvement Award under Grant Number 2003248. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The TopoWx dataset is provided by the Network for Sustainable Climate Risk Management, University Park, PA, USA, from their website at https://www.scrim.psu.edu/resources/topowx/. The GAGES-II dataset is available from https://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml. The PRISM precipitation dataset is available from https://prism.oregonstate.edu/. The streamflow data is available from https://waterwatch.usgs.gov/.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2021.126621.

References

Abatzoglou, J.T., Ficklin, D.L., 2017. Climatic and physiographic controls of spatial variability in surface water balance over the contiguous United States using the Budyko relationship. Water Resour. Res. 53 (9), 7630–7643. https://doi.org/10.1002/2017WR020843.
Bai, P., Liu, X., Zhang, D., Liu, C., 2020. Estimation of the Budyko model parameter for small basins in China. Hydrol. Process. 34 (1), 125–138. https://doi.org/10.1002/hyp.13577.
Bao, Z., Zhang, J., Wang, G., Chen, Q., Guan, T., Yan, X., et al., 2019. The impact of climate variability and land use/cover change on the water balance in the Middle Yellow River Basin. China. J. Hydrol. 577, 123942.
Berghuijs, W.R., Woods, R.A., Hrachowitz, M., 2014. A precipitation shift from snow towards rain leads to a decrease in streamflow. Nat. Clim. Change 4 (7), 583–586. https://doi.org/10.1038/nclimate2246.
Berghuijs, W.R., Larsen, J.R., Van Emmerik, T.H., Woods, R.A., 2017. A global assessment of runoff sensitivity to changes in precipitation, potential evaporation, and other factors. Water Resour. Res. 53 (10), 8475–8486.
Bhaskar, A. S., Hopkins, K. G., Smith, B. K., Stephens, T. A., & Miller, A. J. (2020). Hydrologic Signals and Surprises in U.S. Streamflow Records During Urbanization. Water Resources Research, 56(9), e2019WR027039. https://doi.org/10.1029/2019WR027039.
Bogaert, J., Rousseau, R., Hecke, P.V., Impens, I., 2000. Alternative area-perimeter ratios for measurement of 2D shape compactness of habitats. Appl. Math. Comput. 111 (1), 71–85. https://doi.org/10.1016/S0096-3003(99)00075-2.
Bring, J., 1994. How to standardize regression coefficients. Am. Stat. 48 (3), 209–213.
Chang, H., Johnson, G., Hinkley, T., Jung, I.-W., 2014. Spatial analysis of annual runoff ratios and their variability across the contiguous US. J. Hydrol. 511, 387–402.
Chen, X., Alimohammadi, N., Wang, D., 2013. Modeling interannual variability of seasonal evaporation and storage change based on the extended Budyko framework. Water Resour. Res. 49 (9), 6067–6078. https://doi.org/10.1002/wrcr.20493.
Chen, Z., Wang, W., Woods, R.A., Shao, Q., 2020. Hydrological effects of change in vegetation components across global catchments. J. Hydrol. 125775 https://doi.org/10.1016/j.jhydrol.2020.125775.
Dai, A., 2008. Temperature and pressure dependence of the rain-snow phase transition over land and ocean. Geophys. Res. Lett. 35 (12).
Debbage, N., Shepherd, J.M., 2018. The Influence of Urban Development Patterns on Streamflow Characteristics in the Charlanta Megaregion. Water Resour. Res. 54 (5), 3728–3747. https://doi.org/10.1029/2017WR021594.
Destouni, G., Jaramillo, F., Prieto, C., 2013. Hydroclimatic shifts driven by human water use for food and energy production. Nat. Clim. Change 3 (3), 213–217.
Dey, P., Mishra, A., 2017. Separating the impacts of climate change and human activities on streamflow: A review of methodologies and critical assumptions. J. Hydrol. 548, 278–290. https://doi.org/10.1016/j.jhydrol.2017.03.014.
Dingman, S.L., 2015. Physical hydrology. Waveland press.
Donohue, R.J., Roderick, M.L., McVicar, T.R., 2007. On the importance of including vegetation dynamics in Budyko's hydrological model. Hydrol. Earth Syst. Sci. 11 (2), 983–995. https://doi.org/10.5194/hess-11-983-2007.
Donohue, R.J., Roderick, M.L., McVicar, T.R., 2010. Can dynamic vegetation information improve the accuracy of Budyko's hydrological model? J. Hydrol. 390 (1), 23–34. https://doi.org/10.1016/j.jhydrol.2010.06.025.
Donohue, R.J., Roderick, M.L., McVicar, T.R., 2012. Roots, storms and soil pores: Incorporating key ecohydrological processes into Budyko's hydrological model. J. Hydrol. 436–437, 35–50. https://doi.org/10.1016/j.jhydrol.2012.02.033.
Falcone, J.A., Carlisle, D.M., Weber, L.C., 2010b. Quantifying human disturbance in watersheds: variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. Ecol. Ind. 10 (2), 264–273.

Falcone, J. A., Carlisle, D. M., Wolock, D. M., & Meador, M. R. (2010). GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States: Ecological archives E091-045. Ecology, 91(2), 621–621.
Falcone, J.A., 2011, GAGES–II—Geospatial attributes of gages for evaluating streamflow: U.S. Geological Survey metadata. http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml.
Fang, K., Shen, C., Fisher, J.B., Niu, J., 2016. Improving Budyko curve-based estimates of long-term water partitioning using hydrologic signatures from GRACE. Water Resour. Res. 52 (7), 5537–5554.
Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., et al., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J. Hydrol. 537, 45–60.
Fu, B., 1981. On the calculation of the evaporation from land surface (in Chinese). Scientia Atmospherica Sinica 5, 23–31.
Gan, G., Liu, Y., Sun, G., 2020. Understanding interactions among climate, water, and vegetation with the Budyko framework. Earth Sci. Rev. 103451.
Gentine, P., D'Odorico, P., Lintner, B.R., Sivandran, G., Salvucci, G., 2012. Interdependence of climate, soil, and vegetation as constrained by the Budyko curve. Geophys. Res. Lett. 39 (19).
Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2013). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. ArXiv Preprint ArXiv:1306.0413.
Greve, P., Gudmundsson, L., Orlowsky, B., & Seneviratne, S. I. (2015). Introducing a probabilistic Budyko framework. Geophysical Research Letters, 42(7), 2261–2269.
Greve, P., Burek, P., & Wada, Y. (2020). Using the Budyko framework for calibrating a global hydrological model. Water Resources Research, 56(6), e2019WR026280.
Gudmundsson, L., Greve, P., Seneviratne, S.I., 2016. The sensitivity of water availability to changes in the aridity index and other factors—A probabilistic analysis in the Budyko space. Geophys. Res. Lett. 43 (13), 6985–6994.
Han, S., Hu, H., Yang, D., Liu, Q., 2011. Irrigation impact on annual water balance of the oases in Tarim Basin, Northwest China. Hydrol. Process. 25 (2), 167–174. https://doi.org/10.1002/hyp.7830.
Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. J. Irrig. Drainage Div. 108 (3), 225–230.
Hodgkins, G.A., Dudley, R.W., Archfield, S.A., Renard, B., 2019. Effects of climate, regulation, and urbanization on historical flood trends in the United States. J. Hydrol. 573, 697–709.
Huo, J., Liu, C., Yu, X., Jia, G., Chen, L., 2021. Effects of watershed char and climate variables on annual runoff in different climatic zones in China. Sci. Total Environ. 754, 142157 https://doi.org/10.1016/j.scitotenv.2020.142157.
Jiang, C., Xiong, L., Wang, D., Liu, P., Guo, S., Xu, C.-Y., 2015. Separating the impacts of climate change and human activities on runoff using the Budyko-type equations with time-varying parameters. J. Hydrol. 522, 326–338. https://doi.org/10.1016/j.jhydrol.2014.12.060.
Jung, K., Marpu, P.R., Ouarda, T.B., 2017. Impact of river network type on the time of concentration. Arab. J. Geosci. 10 (24), 546.
Kelly, S.A., Takbiri, Z., Belmont, P., Foufoula-Georgiou, E., 2016. Human amplified changes in precipitation-runoff patterns in large river basins of the Midwestern United States. Hydrol. Earth Syst. Sci. 1.
Kopecký, M., Čížková, Š., 2010. Using topographic wetness index in vegetation ecology: does the algorithm matter? Appl. Veg. Sci. 13 (4), 450–459. https://doi.org/10.1111/j.1654-109X.2010.01083.x.
Li, D., Pan, M., Cong, Z., Zhang, L., Wood, E., 2013. Vegetation control on water and energy balance within the Budyko framework. Water Resour. Res. 49 (2), 969–976. https://doi.org/10.1002/wrcr.20107.
Li, C., Sun, G., Caldwell, P.V., Cohen, E., Fang, Y., Zhang, Y., et al., 2020. Impacts of Urbanization on Watershed Water Balances across the Conterminous United States. Water Resour. Res. e2019WR026574.
Liang, W., Bai, D., Wang, F., Fu, B., Yan, J., Wang, S., et al., 2015. Quantifying the impacts of climate change and ecological restoration on streamflow changes based on a B udyko hydrological model in C hina's L oess P lateau. Water Resour. Res. 51 (8), 6500–6519.
Magilligan, F.J., Nislow, K.H., 2005. Changes in hydrologic regime by dams. Geomorphology 71 (1), 61–78. https://doi.org/10.1016/j.geomorph.2004.08.017.
Markham, C.G., 1970. Seasonality of precipitation in the United States. Ann. Assoc. Am. Geogr. 60 (3), 593–597.
Maupin, M. A., Kenny, J. F., Hutson, S. S., Lovelace, J. K., Barber, N. L., & Linsey, K. S. (2014). Estimated use of water in the United States in 2010. US Geological Survey.
Menard, S., 2002. Applied logistic regression analysis, Vol. 106. Sage.
Milly, P.C.D., 1994. Climate, soil water storage, and the average annual water balance. Water Resour. Res. 30 (7), 2143–2156. https://doi.org/10.1029/94WR00586.
Milly, P.C.D., Dunne, K.A., 2020. Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation. Science 367 (6483), 1252–1255. https://doi.org/10.1126/science.aay9187.
Mohammadinia, A., Alimohammadi, A., Saeidian, B., 2017. Efficiency of geographically weighted regression in modeling human leptospirosis based on environmental factors in Gilan Province, Iran. Geosciences 7 (4), 136.
Nayak, A.K., Biswal, B., Sudheer, K.P., 2020. A novel framework to determine the usefulness of satellite-based soil moisture data in streamflow prediction using Dynamic Budyko model. J. Hydrol. 125849.
Neto, A.A.M., Niu, G.-Y., Roy, T., Tyler, S., Troch, P.A., 2020. Interactions between snow cover and evaporation lead to higher sensitivity of streamflow to temperature. Commun. Earth Environ 1 (1), 1–7. https://doi.org/10.1038/s43247-020-00056-9.
Ning, T., Zhou, S., Chang, F., Shen, H., Li, Z., Liu, W., 2019. Interaction of vegetation, climate and topography on evapotranspiration modelling at different time scales within the Budyko framework. Agric. For. Meteorol. 275, 59–68. https://doi.org/10.1016/j.agrformet.2019.05.001.
Oliveira, K.D., Tomasella, J., Sanches, L.D., 2019. Spatial-temporal analysis of the climatic and anthropogenic influences on runoff in the Jucu River Basin, Southeastern Brazil. Land Degrad. Develop. 30 (17), 2073–2087. https://doi.org/10.1002/ldr.3403.
Olivera, F., DeFee, B.B., 2007. Urbanization and its effect on runoff in the whiteoak bayou watershed, Texas1. JAWRA J. Am. Water Resour. Assoc. 43 (1), 170–182. https://doi.org/10.1111/j.1752-1688.2007.00014.x.
Oudin, L., Salavati, B., Furusho-Percot, C., Ribstein, P., Saadi, M., 2018. Hydrological impacts of urbanization at the catchment scale. J. Hydrol. 559, 774–786.
Oyler, J.W., Ballantyne, A., Jencso, K., Sweet, M., Running, S.W., 2015. Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. Int. J. Climatol. 35 (9), 2258–2279.
Padrón, R.S., Gudmundsson, L., Greve, P., Seneviratne, S.I., 2017. Large-scale controls of the surface water balance over land: Insights from a systematic review and meta-analysis. Water Resour. Res. 53 (11), 9659–9678.
Patterson, L.A., Lutz, B., Doyle, M.W., 2013. Climate and direct human contributions to changes in mean annual streamflow in the South Atlantic, USA. Water Resour. Res. 49 (11), 7278–7291. https://doi.org/10.1002/2013WR014618.
Pike, R.J., Wilson, S.E., 1971. Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. Geol. Soc. Am. Bull. 82 (4), 1079–1084.
Potter, N.J., Zhang, L., Milly, P.C.D., McMahon, T.A., Jakeman, A.J., 2005. Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments. Water Resour. Res. 41 (6) https://doi.org/10.1029/2004WR003697.
PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 18 August 2014.
Renner, M., Seppelt, R., Bernhofer, C., Schymanski, S., 2012. Evaluation of water-energy balance frameworks to predict the sensitivity of streamflow to climate change. Hydrol. Earth Syst. Sci. 16 (5).
Rennermalm, A.K., Bring, A., Mote, T.L., 2012. Spatial and scale-dependent controls on North American Pan-Arctic minimum river discharge. Geograph. Anal. 44 (3), 202–218.
Rice, J.S., Emanuel, R.E., Vose, J.M., Nelson, S.A.C., 2015. Continental U.S. streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics. Water Resour. Res. 51 (8), 6262–6275. https://doi.org/10.1002/2014WR016367.
Rodell, M., Famiglietti, J.S., Wiese, D.N., Reager, J.T., Beaudoing, H.K., Landerer, F.W., Lo, M.-H., 2018. Emerging trends in global freshwater availability. Nature 557 (7707), 651.
Sankarasubramanian, A., Vogel, R.M., Limbrunner, J.F., 2001. Climate elasticity of streamflow in the United States. Water Resour. Res. 37 (6), 1771–1781.
Sawicz, K.A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., Carrillo, G., 2014. Characterizing hydrologic change through catchment classification. Hydrol. Earth Syst. Sci. 18 (1), 273–285. https://doi.org/10.5194/hess-18-273-2014.
Shao, Q., Traylen, A., Zhang, L., 2012. Nonparametric method for estimating the effects of climatic and catchment characteristics on mean annual evapotranspiration. Water Resour. Res. 48 (3) https://doi.org/10.1029/2010WR009610.
Sinha, J., Jha, S., Goyal, M.K., 2019. Influences of watershed characteristics on long-term annual and intra-annual water balances over India. J. Hydrol. 577, 123970.
Wada, Y., Wisser, D., Bierkens, M.F., 2014. Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth Syst. Dyn. 5 (1), 15–40.
Walsh, R.P.D., Lawler, D.M., 1981. Rainfall seasonality: description, spatial patterns and change through time. Weather 36 (7), 201–208.
Wang, D., Hejazi, M., 2011. Quantifying the relative contribution of the climate and direct human impacts on mean annual streamflow in the contiguous United States. Water Resour. Res. 47 (10) https://doi.org/10.1029/2010WR010283.
Wang, W., Zhang, Y., Tang, Q., 2020. Impact assessment of climate change and human activities on streamflow signatures in the Yellow River Basin using the Budyko hypothesis and derived differential equation. J. Hydrol. 591, 125460.
Wheeler, D.C., 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environ. Plan. A 39 (10), 2464–2481.
Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geogr. Syst. 7 (2), 161–187.
Wheeler, D. C. (2019). Geographically Weighted Regression. In M. M. Fischer & P. Nijkamp (Eds.), Handbook of Regional Science (pp. 1–27). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36203-3_77-1.
Wickham, J.D., Stehman, S.V., Gass, L., Dewitz, J., Fry, J.A., Wade, T.G., 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. Remote Sens. Environ. 130, 294–304.
Wolock, D.M., McCabe Jr., G.J., 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. Water Resour. Res. 31 (5), 1315–1324. https://doi.org/10.1029/95WR00471.
Wolock, D.M., McCabe, G.J., 1999. Explaining spatial variability in mean annual runoff in the conterminous United States. Climate Res. 11 (2), 11.
Xing, W., Wang, W., Shao, Q., Yong, B., 2018. Identification of dominant interactions between climatic seasonality, catchment characteristics and agricultural activities on Budyko-type equation parameter estimation. J. Hydrol. 556, 585–599.
Xu, X., Liu, W., Scanlon, B.R., Zhang, L., Pan, M., 2013. Local and global factors controlling water-energy balances within the Budyko framework. Geophys. Res. Lett. 40 (23), 6123–6129. https://doi.org/10.1002/2013GL058324.

Yacim, J.A., Boshoff, D.G.B., 2019. A Comparison of bandwidth and kernel function selection in geographically weighted regression for house valuation. Int. J. Technol. 10 (1), 58–68.

Yang, D., Sun, F., Liu, Z., Cong, Z., Ni, G., Lei, Z., 2007. Analyzing spatial and temporal variability of annual water-energy balance in nonhumid regions of China using the Budyko hypothesis. Water Resour. Res. 43 (4) https://doi.org/10.1029/2006WR005224.

Yang, D., Shao, W., Yeh, P.-J.-F., Yang, H., Kanae, S., Oki, T., 2009. Impact of vegetation coverage on regional water balance in the nonhumid regions of China. Water Resour. Res. 45 (7) https://doi.org/10.1029/2008WR006948.

Yang, H., Yang, D., Lei, Z., Lei, H., 2008a. Derivation and validation of watershed coupled water-energy balance equation at arbitrary time scale (in Chinese). J. Hydraulic 39, 610–617.

Yang, H., Yang, D., Lei, Z., Sun, F., 2008b. New analytical derivation of the mean annual water-energy balance equation. Water Resour. Res. 44 (3) https://doi.org/10.1029/2007WR006135.

Zhang, X., Dong, Q., Cheng, L., Xia, J., 2019. A Budyko-based framework for quantifying the impacts of aridity index and other factors on annual runoff. J. Hydrol. 579, 124224.

Zhang, L., Hickel, K., Dawes, W.R., Chiew, F.H.S., Western, A.W., Briggs, P.R., 2004. A rational function approach for estimating mean annual evapotranspiration. Water Resour. Res. 40 (2) https://doi.org/10.1029/2003WR002710.

Zhang, S., Yang, H., Yang, D., Jayawardena, A.W., 2016. Quantifying the effect of vegetation change on the regional water balance within the Budyko framework. Geophys. Res. Lett. 43 (3), 1140–1148. https://doi.org/10.1002/2015GL066952.